UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

ENAI TAVEIRA DA CUNHA

On INAR(1) models for integer time series

Recife
2018

ENAI TAVEIRA DA CUNHA

ON INAR(1) MODELS FOR INTEGER TIME SERIES

Doctoral thesis submitted to the Graduate Program in Statistics, Department of Statistics, Federal University of Pernambuco as a partial requirement for obtaining a doctorate in Statistics.

Advisor: Prof. Dr. Klaus Leite Pinto Vasconcellos
Co-Advisor: Prof. Dr. Marcelo Bourguignon Pereira

Recife
2018

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

C972a Cunha, Enai Taveira da
On INAR (1) models for integer time series / Enai Taveira da Cunha. – 2018.
89 f.: il., fig., tab.

Orientador: Klaus Leite Pinto Vasconcellos.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2018.
Inclui referências.

1. Estatística. 2. Correção de viés. I. Vasconcellos, Klaus Leite Pinto (orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2018-037

ENAI TAVEIRA DA CUNHA

ON INAR (1) MODELS FOR INTEGER TIME SERIES

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Estatística.

Aprovada em: 22 de fevereiro de 2018.

BANCA EXAMINADORA

Prof. Klaus Leite Pinto Vasconcellos
UFPE

Prof. Francisco Cribari Neto
UFPE

Prof. Abraão David Costa do Nascimento
UFPE

Prof. Francisco Sousa Ramos
UFPE

Prof.ª Luz Milena Zéa Fernandez
UFRN

Dedico este trabalho a minha mãe e protetora Maria José.

Agradecimentos

Quero agradecer em primeiro lugar a Deus, por me guiar durante todo meu caminho, por me dar paciência e sabedoria em cada momento de dificuldade e por nunca me deixar desistir.

A minha mãe, responsável pela minha vida, mulher guerreira, sábia e principal motivo para eu nunca desistir dos meus objetivos.

Aos meus irmãos, Janaína, Janilson e Neto. Obrigada por toda compreensão e colaboração para que tudo desse certo.

Ao meu orientador, Professor Dr. Klaus Leite Pinto Vasconcellos, por toda paciência que teve comigo durante todo esse processo de doutoramento, pela sua disposição, por todos os ensinamentos e pelas contribuições que possibilitaram a concretização desse trabalho.

Ao meu coorientador, Professor Dr. Marcelo Bourguignon Pereira, por todo o acompanhamento que me deu durante o desenvolvimento do trabalho, por sua paciência, apoio e dedicação. Por sua contribuição que tornou esse trabalho possível.

Aos professores do Programa de Pós-Graduação em Estatística, pelos ensinamentos e dedicação nas disciplinas.

Aos meus amigos da UFPE, com destaque para os amigos Cláudio, Cristiane, Dany, Jéssica, Giannini, Josimar, Pedro Rafael, Renilma, Thiago Vedovato e Wênia. Obrigada por estarem presentes nos momentos bons e nos momentos difíceis. A amizade de vocês foi essencial para completar essa jornada.

Às minhas amigas Iolanda e Karine, por toda a convivência que tivemos, toda amizade e cumplicidade.

À Universidade Federal Rural do Semi-Árido pelo afastamento concedido para minha capacitação.

A Valéria Bittencourt, secretária da pós-graduação em estatística, pela competência, dedicação, amizade, carinho e atenção.

Ao meu noivo Ramon Jales, por estar sempre ao meu lado, me dando apoio, carinho e atenção.

Três coisas são necessárias para se carregar os pesos da vida:
A esperança, o sono e o riso.
(Immanuel Kant).

Abstract

Modelling counts of events can be found in several situations of real life. For instance, the number of customers in a department store per day, monthly number of cases of some disease or the number of thunderstorms in a day. The study of integer-valued time series has grown greatly in recent decades, the reason for this is the need of appropriate models for the statistical analysis of count time series. Motivated for this, the topic of this work is integer-valued time series models. This thesis is divided into three parts, composed by three independent papers about integer-valued time series models. A brief review of the three chapters can be seen below. The skew integer-valued time series process with generalized Poisson difference distribution marginal is introduced in Chapter 2. A new thinning operator is defined as the difference of two quasi-binomial thinning operators and the new process is defined based on it. Some properties of the process like mean, variance, skewness and kurtosis are presented. The conditional expectation and variance are obtained, the autocorrelation and spectral function are derived. The moments estimation is considered and a Monte Carlo simulation is presented to study a performance of moments estimators. An application to a real data set is discussed. In Chapter 3, we consider the first-order integer-valued autoregressive process with geometric marginal distributions, NGINAR(1) process, and develop a nearly unbiased estimator for one of the parameters of the process. We consider the Yule-Walker estimators, derive the first order bias for one of the parameters and propose a new bias-adjusted estimator. Monte Carlo simulation studies are considered to analyse the behaviour of the new estimator. Finally, in Chapter 4 we introduce a first order integer-valued autoregressive process with Borel innovations based on the binomial thinning operator. This model is suitable to modelling zero truncated count time series with equidispersion, underdispersion and overdispersion. The basic properties of the process are obtained. To estimate the unknown parameters, the Yule-Walker, conditional least squares and conditional maximum likelihood methods are considered. The asymptotic distribution of conditional least squares estimators is obtained and hypothesis tests for an equidispersed model against an underdispersed or overdispersed model are formulated. A Monte Carlo simulation is presented analysing the estimators performance in finite samples. Two applications to real data are presented to show that the Borel INAR(1) model is suited to model underdispersed and overdispersed data counts.

Keywords: Bias correction. Borel distribution. Conditional least squares estimator. Generalized Poisson distribution. INAR(1) process. Yule-Walker estimator.

Resumo

Eventos de contagem podem ser encontrados em muitas situações práticas. Por exemplo, o número de clientes em uma loja de departamentos em um dia, o número mensal de casos de alguma doença ou o número de tempestades em um dia. O estudo de séries temporais assumindo valores inteiros cresceu muito nas recentes décadas, a razão para isto é a necessidade de modelos apropriados para a análise estatística de séries temporais de contagem. Motivado por isto, o tópico principal deste trabalho é modelos de séries temporais com valores inteiros. Esta tese é composta por três artigos, todos dentro desta área e cada capítulo pode ser lido independentemente um do outro. No segundo capítulo, definimos um novo processo autorregressivo para valores inteiros com distribuição marginal diferença de Poisson generalizadas, baseado na diferença de dois operadores "thinning" quase-binomiais. Este modelo é adequado para conjuntos de dados com valores positivos e negativos e pode ser visto como uma generalização do processo INAR(1) com marginal diferença de Poisson. Uma das vantagens da diferença de duas variáveis aleatórias com distribuição Poisson generalizada é que, em comparação com a diferença de variáveis aleatórias com distribuição Poisson, esta pode apresentar cauda curta ou cauda longa, sendo mais flexível aos dados. Algumas propriedades estatísticas básicas do processo e da distribuição condicional são obtidas. Os estimadores de Yule-Walker são considerados para a estimação dos parâmetros desconhecidos do modelo e simulações de Monte Carlo são apresentadas para o estudo da performance dos estimadores. Uma aplicação a um conjunto de dados reais é discutida para mostrar o potencial do modelo na prática. No capítulo seguinte, desenvolvemos estimadores aproximadamente não viesados para um dos parâmetros do processo autoregressivo de valores inteiros de primeira ordem com distribuição marginal geométrica, processo NGINAR(1). Consideramos os estimadores de Yule-Walker do processo, derivamos o viés de primeira ordem para um dos parâmetros do modelo e propomos um novo estimador com viés ajustado. Um estudo de simulação de Monte Carlo é considerado para analisar o comportamento do novo estimador. Finalmente, no capítulo 4 introduzimos o processo autorregressivo de primeira ordem com inovações Borel baseado no "thinning" binomial. Este modelo é adequado para séries temporais de contagem truncadas no zero que apresentam equidispersão, subdispersão e sobredispersão. Propriedades básicas do processo são obtidas. Para estimar os parâmetros desconhecidos são considerados os métodos de estimação de Yule-Walker, mínimos quadrados condicionais e máxima verossimilhança condicional. A distribuição assintótica dos estimadores de mínimos quadrados condicionais é obtida e testes de hipótese para testar um modelo equidisperso contra um subdisperso ou sobredisperso são formulados. Simulação de Monte Carlo é apresentada, analisando a performance dos estimadores em amostras finitas.

Duas aplicações com dados reais são apresentadas mostrando que o modelo Borel INAR(1) é adequado para dados de contagens com subdispersão e sobredispersão.

Palavras-chave: Correção de viés. Distribuição de Borel. Distribuição Poisson generalizada. Estimador de mínimos quadrados condicional. Estimador de Yule-Walker. Processos INAR(1).

List of Figures

List of Tables

Contents

1 Introduction

Integer-valued time series is a topic that has received much attention in the last three decades. The reason for this is the need of appropriate models for the statistical analysis of count time series. Count time series occur naturally in several fields. In medicine, the number of patients in a hospital at a specific point of time or the number of epileptic seizures a patient suffers each day. In economics, the number of insurance claims a company receives during each day or discrete transaction price movements on financial market. Additionally, the number of traffic accidents, the number of crime ocurrences and so on.

Pioneering works in this theme were McKenzie (1985), Al-Osh and Alzaid (1987) and McKenzie (1988). These authors introduced in literature the first-order non-negative integer-valued autoregressive (INAR(1)) process based on the binomial thinning operator by Steutel and Van Harn (1979).

The idea of thinning operation is to obtain an operator that works in discrete time series in the same way the multiplication in the classical autoregressive (AR) processes (Box et al, 1994). Since the procedure of multiplying an integer-valued random variable by a real constant does not necessarily leads to an integer valued random variable, several authors defined integer-valued time series making use of thinning operations. Weiß (2008) reviewed some thinning operations and showed how they are successfully applied to define integer-valued ARMA models.

Several models for count data with different marginal distributions were defined based on different thinning operations. For instance, McKenzie (1985) reviews integer-valued time-series models with Poisson, geometric, negative-binomial and binomial marginal distributions. McKenzie (1986) defined the autoregressive moving-average process with negative-binomial and geometric marginal distributions. McKenzie (1988) developed a family of models for discrete-time processes with Poisson marginal distributions. An extension of INAR(1) process, the integer-valued pth-order autoregressive structure (INAR(p)) process was considered by Alzaid and Al-Osh (1990) and Du and Li (1991). Al-Osh and Alzaid (1991) introduced a family of models for a stationary sequence of dependent binomial random variables. Alzaid and Al-Osh (1993) defined an autoregressive integer-valued time series with generalized Poisson marginal distribution. A first-order integer-valued autoregressive process with geometric marginal distribution (NGINAR) was introduced by Ristić et al. (2009).

Besides the formulation of the new models, issues as properties (Bakouch, 2010; Weiß and Kim, 2011), estimation (Jung et al. 2005) and asymptotic distributions of model estimators (Freeland and McCabe, 2005) have been considered.

This thesis is divided into three parts, composed by three independent papers. So, we decided that, for this thesis, each of the papers fills a distinct chapter. Therefore, each chapter can be read independently of each other, since each is self-contained. Additionally, we emphasize that each chapter contains a thorough introduction to the presented matter, so this general introduction only shows, quite briefly, the context of each chapter. A brief review of the three chapters can be seen below.

The skew integer-valued time series process with generalized Poisson difference marginal distribution is introduced in Chapter 2. A new thinning operator is defined based on difference of two quasi-binomial thinning operators and the new process is defined based on it similarly as Freeland (2010). Some properties of the process like mean, variance, skewness and kurtosis are presented. The conditional expectation and the variance are obtained, the autocorrelation and spectral functions are derived. The moments estimation is considered and a Monte Carlo simulation study is presented to quantify a performance of moments estimators. An application to real data set is discussed.

In Chapter 3, we developed nearly unbiased estimators for the parameters of the first-order integer-valued autoregressive process with geometric marginal distributions (NGINAR(1)) based on negative binomial thinning (Ristić et al. (2009)). We consider the Yule-Walker estimators, derive the first order bias for one of the parameters and propose a new bias-adjusted estimator. A Monte Carlo simulation study is considered to analyse the behaviour of the new estimator.

Chapter 4 develops a new time series model based on the Borel distribution (Consul and Famoye, 2006). We introduce the shifted first order integer-valued autoregressive process with Borel innovations. This model is defined based on the binomial thinning with Borel distribution innovations. The Borel INAR(1) model is suited to equidispersed, underdispersed and overdispersed data counts set. The basic properties of this model are obtained. Closed form expressions for the conditional least square (CLS) and Yule-Walker estimators of unknown parameters are derived and we obtain their asymptotic distributions. A hypothesis test based on the asymptotic distribution of CLS estimator is defined for equidispersed against overdispersed or underdispersed model. We perform a test power analysis by simulating the rejection rate for different sample sizes considering the true and estimated value of α and the asymptotic distribution. Two applications are presented showing that the Borel INAR(1) model is suited to shifted underdispersed and overdispersed data sets and the results of a hypothesis test confirm the assessment.

2 A skew integer-valued time series process with generalized Poisson difference marginal distribution

Resumo

Neste capítulo, propomos um novo processo autorregressivo para valores inteiros com distribuição marginal diferença de Poisson generalizada baseado na diferença de dois operadores "thinning" quase-binomial. Este modelo é adequado para conjuntos de dados sob $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ e pode ser visto como uma generalização do processo INAR(1) com marginal diferença de Poisson. Uma das vantagens da diferença de duas variáveis aleatórias com distribuição Poisson generalizada é que, comparada com a diferença de variáveis aleatórias com distribuição Poisson, pode apresentar cauda curta ou cauda longa, sendo mais flexível aos dados. Algumas propriedades básicas do processo como média, variancia, coeficiente de assimetria, curtose e outras propriedades condicionais são obtidas. Os estimadores de Yule-Walker são considerados para a estimação dos parâmetros desconhecidos do modelo e simulações de Monte Carlo são apresentadas para o estudo da performance dos estimadores. Uma aplicação a um conjunto de dados reais é feita para ilustrar o modelo proposto.

Palavras-chave: Distribuição quase-binomial. Processo INAR(1). Processo Poisson generalizada. Séries temporais de valores inteiros.

Abstract

In this chapter, we propose a new integer-valued autoregressive process with generalized Poisson difference marginal distributions based on difference of two quasi-binomial thinning operators. This model is suitable for data sets on $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ and can be viewed as a generalisation of the Poisson difference INAR(1) process. An advantage of the difference of two generalized Poisson random variables is it can have longer or shorter tails compared to the Poisson difference distribution. We present some basic properties of our process like mean, variance, skewness, kurtosis and other conditional properties of the process are derived. To estimate unknown parameters of the model, the Yule-Walker estimators are considered and a

Monte Carlo simulation is made to study the performance of estimators. An application to a real data set is discussed to show the potential for practice of our model.

Keywords: Generalized Poisson process. Quasi-binomial distribution. INAR(1) process. Integer-valued time series.

2.1 Introduction

In the last three decades several authors have discussed models for integer-valued time series. Al-Osh and Alzaid (1987) introduced the first-order integer-valued autoregressive (INAR(1)) process with Poisson marginal using the binomial thinning operator introduced by Steutal and Van Harn (1979). Alzaid and Al-Osh (1990) generalized the INAR process, defining an integer-valued pth-order autoregressive structure, INAR(p) process, in the same way of INAR(1) process. Du and Li (1991) proposed the integer-valued autoregressive INAR(p) model using the binomial thinning operator and added the hypotheses that the innovations series independent of the counting series. A first-order autoregressive process with generalized Poisson marginal distribution (GPAR(1)) was defined by Alzaid and Al-Osh (1993) based on the quasi-binomial thinning operator. Ristić et al. (2009) introduced a count valued time series model with geometric marginal with negative binomial thinning operator. This model is an alternative to the Poisson INAR(1) when the data set is overdispersed. Several other models were defined for time series with non-negative values with different thinning operators (see Weiß, 2008). In many fields it is common to have data sets based on negative and positive values, such as medicine (Karlis and Ntzoufras, 2006), financial studies (Shahtahmassebi and Moyeed, 2013, Chesneau and Kachour, 2012), sports (Shahtahmassebi and Moyeed, 2016). Thus, it becomes necessary to define integer-valued time series models for such specific data sets. Recently, many authors have presented time series models when the positive and negative integer values arise as the difference between two discrete distributions. Freeland (2010) defined a first-order stationary autoregressive process on $\mathbb{Z}$ with symmetric Skellam marginals, the new process was defined by a difference between two independent and identically distributed random variables with Poisson distribution and a new thinning operator obtained by the difference of two binomial thinning operators, the process was called True INAR(1) (TINAR(1)). Barreto-Souza and Bourguignon (2015) defined a modified negative binomial thinning operator by the difference of two negative binomial thinning operators and defined an INAR(1) process on $\mathbb{Z}$ with skew discrete Laplace marginals by the difference of two independent NGINAR(1) (Ristić et al., 2009) process with geometric marginals with means $\mu_1 > 0$ and $\mu_2 > 0$, respectively. Some properties of the process were derived, including joint and conditional basic properties, characteristic function, moments and estimators for the parameters. Another thinning operator using the difference of two negative binomial thinning operators was defined by Nástic et al. (2016). Based on the difference of two independent random variables with geometric marginal distributions with the same mean $\mu > 0$, they defined a new symmetric INAR(1) process with discrete Laplace marginal distributions with positive or negative lag-one autocorrelation. The properties of the new thinning operator were derived and properties of the process were obtained. Alzaid and Omair (2014) defined a new operator called the extended binomial thinning operator and introduced the Poisson difference integer valued autoregressive model of order one, PDINAR(1) process, with Poisson difference marginal, Skellam distribution, to model integer valued time series with possible negative values and either positive or negative correlations. This model was applied to data from the Saudi stock exchange. A stationary first-order integer valued autoregressive process with geometric-Poisson marginals (NSINAR(1)) was introduced by Bourguingnon and Vasconcellos (2016). The new process allows negative values in the series. Several properties were established and estimators are defined for unknown parameters by the Yule-Walker method. This process was defined as the difference between geometric and Poisson random variables and the new thinning operator based on the difference of the negative binomial thinning operator and the binomial thinning operator.

In this chapter, we propose a new thinning operator based on the idea of Freeland (2010), using the sequence of quasi-binomial thinning operator defined by Alzaid and Al-Osh (1993). Then we define the skew integer valued autoregressive process with generalized Poisson difference marginal distribution (GPDAR(1)) as the difference between two random variables that follow generalized Poisson (GP) distribution. The motivation for such a process arises from the necessity to have adequate models that can be adapted to skew integer valued (including both negative and positive values) time series. Moreover, this process with generalized Poisson difference (GPD) marginal has more flexibility in the tails than the process with Poisson difference (PD) marginal distribution, that, in the some cases, has a tendency to underestimate values in the tails (see Shahtahmassebi and Moyeed, 2014). The distribution of the difference of two GP random variables can have longer or shorter tails compared to the PD distribution and the GPD distribution can be viewed as a generalisation of the PD distribution.

The chapter is organized as follow: In Section 2, we present a review of the GP and quasi-binomial (QB) distributions, and the GPAR(1) process. The GPDAR(1) process is defined based on the difference of two quasi-binomial thinning operators. In Section 3, some of its properties are derived. In Section 4, the Yule-Walker estimators of unknown parameters are presented and numerical results using Monte Carlo simulation are discussed. In Section 5, an application to a real life data set is presented. Finally, we conclude the chapter in Section 6.

2.2 GPDAR(1) process

In this section, we present a stationary first-order integer-valued autoregressive process with generalized Poisson difference marginals. To that end, we first present a brief review of the GP distribution (Consul and Jain, 1973), QB distribution (Consul and Mital, 1975), and the GPAR(1) process.

Consul and Jain (1973) defined a new distribution as a generalization of the Poisson distribution, the GP distribution with two parameters. It was obtained as a limiting form of the generalized negative binomial distribution. Consul (1989) discussed the properties and applications of the GP distribution. For example, the author described the application of GP distribution to the number of chromosome aberrations induced by chemical and physical agents in human and animal cells. Johnson, Kotz and Kemp (1992) have also described some properties of the GP distribution. Joe and Zhu (2005) proved that the GP distribution is a mixture of Poisson distributions and he made a study comparison between the GP and negative binomial distributions. Consul (1986), considering two independent random variables with GP distribution, defined and obtained the properties of the difference between them.

The probability mass function (pmf) of a random variable (r.v.) X following a generalized Poisson distribution with parameters λ and θ, denoted by GP(λ, θ), is given by

$$P(X = x) = \lambda(\lambda + \theta x)^{x-1}\mathrm{e}^{-(\lambda+\theta x)}/x!, \quad x = 0, 1, 2, \ldots, \tag{2.1}$$

where $\lambda > 0$ and $0 \leq \theta < 1$. The GP distribution is also defined for $\theta < 0$, but, in this work, we shall only consider $\theta > 0$. For more details, see Consul and Jain (1973).

Remark 2.2.1. Note that for $\theta = 0$, we obtain the Poisson distribution with parameter λ as a special case of the generalized Poisson distribution.

We can obtain some properties about the GP distribution from Johnson, Kotz and Kemp (2005). We have that the probability generating function (pgf) is given by

$$G(s) = \mathrm{e}^{\lambda(t-1)}, \quad \text{where } t \text{ satisfies } t = s\,\mathrm{e}^{\theta(t-1)},$$

and the mean, variance, the third and fourth central moments are given by

$$\mu = \mathrm{E(X)} = \frac{\lambda}{(1-\theta)}, \quad \sigma^2 = \mathrm{Var(X)} = \frac{\lambda}{(1-\theta)^3}, \tag{2.2}$$

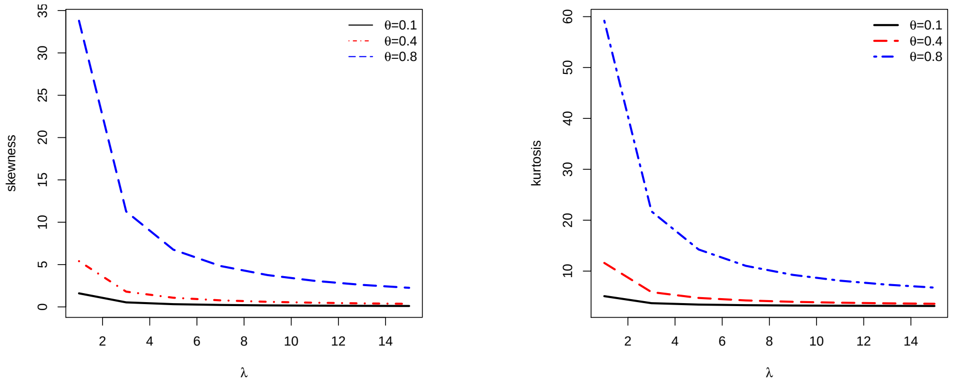
$$\mu_3 = \frac{\lambda(1+2\theta)}{(1-\theta)^5}, \quad \mu_4 = \frac{3\lambda^2}{(1-\theta)^6} + \frac{\lambda(1+8\theta+6\theta^2)}{(1-\theta)^7}.$$

The coefficients of skewness and kurtosis are, respectively,

$$\beta_1 = \frac{(1+2\theta)^2}{\lambda(1-\theta)}, \quad \beta_2 = 3 + \frac{1+8\theta+6\theta^2}{\lambda(1-\theta)}. \tag{2.3}$$

Figure 4.1 displays the skewness and kurtosis of GP distribution for fixed $\theta = 0.1, 0.4, 0.8$ and $\lambda \in \{1, 3, 5, 7, 9, 11, 13, 15\}$.

Note that the skewness of the GP distribution depends on the values of λ for any fixed value of θ. As the value of λ increases, the skewness of GP distribution decreases and for λ enough large the skewness converge to 0. For any fixed value of λ, the skewness becomes large for θ values close to 1.



(a) Skewness of GP distribution (b) Kurtosis of GP distribution

Figure 2.1: Skewness and kurtosis of GP distribution.

Remark 2.2.2. From (2.2) the mean of the GP distribution is smaller than the variance (remember that in this work we are considering $0 \leq \theta < 1$), which means the GP distribution, in this case, may be used as a model to overdispersed data set. Furthermore, from (2.3), the GP distribution adequate to model a skewed data set.

Consul (1986) considered two independent random variables X and Y with $\mathrm{GP}(\lambda_1, \theta)$ and $\mathrm{GP}(\lambda_2, \theta)$ distributions, respectively, with pmf as (2.1), defined a difference distribution $Z \overset{d}{=} X - Y$ and obtained its properties as pmf, cumulants and pgf. The generalized Poisson difference distribution with parameters $\lambda_1, \lambda_2, \theta$ ($\mathrm{GPD}(\lambda_1, \lambda_2, \theta)$) has pmf given by

$$P(Z = z) = \mathrm{e}^{-\lambda_1 - \lambda_2 - z\theta} \sum_{y=0}^{\infty} (\lambda_2, \theta)_y (\lambda_1, \theta)_{y+z} \mathrm{e}^{-2y\theta}, \quad z \in \mathbb{Z},$$

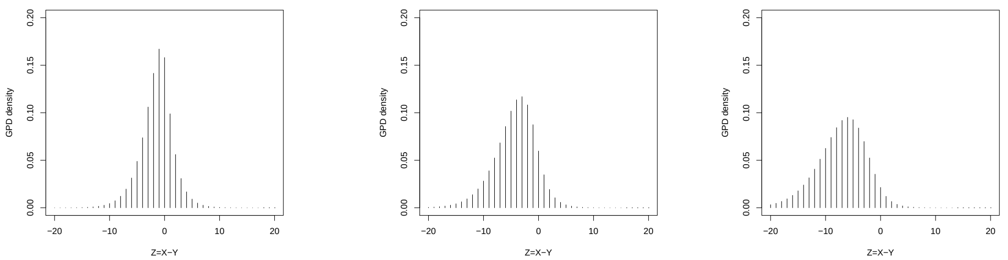
where $(\lambda, \theta)_y = \frac{\lambda(\lambda + \theta y)^{y-1}}{y!}$. Properties of this distribution will be presented in Section 2.3

In Figure 2.2, we can observe that for $\lambda_2 > \lambda_1$ the GPD tends to be left skewed and as the value λ_2 increases the GPD tends to have a longer tail. Note that the following type of symmetry holds to GPD distribution:

$$P(Z = z; \lambda_1, \lambda_2, \theta) = P(Z = -z; \lambda_2, \lambda_1, \theta).$$

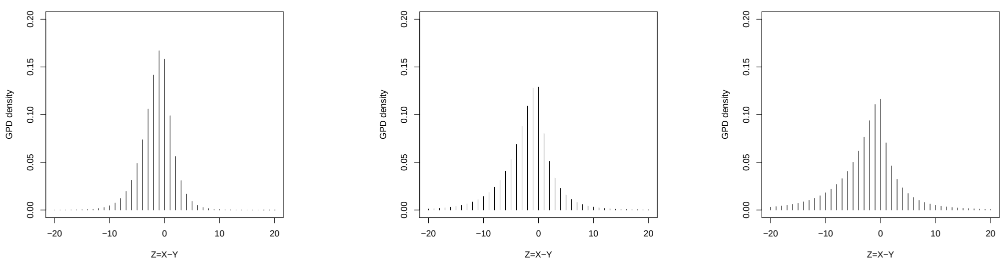
Thus, the GPD behaviour for $\lambda_1 > \lambda_2$ is symmetric to $\lambda_1 < \lambda_2$. In the case of $\lambda_1 = \lambda_2$ we have a symmetric process with zero mean as will see in Section 2.3. Note that, from Figure 2.4 that, fixed θ, when λ_1 and λ_2 increases the tails become heavier.

Finally, observing Figure 2.3 we can see that when θ increases the tails of the GPD distribution become heavier.



(a) $\lambda_1 = 1, \theta = 0.3, \lambda_2 = 2$ (b) $\lambda_1 = 1, \theta = 0.3, \lambda_2 = 4$ (c) $\lambda_1 = 1, \theta = 0.3, \lambda_2 = 6$

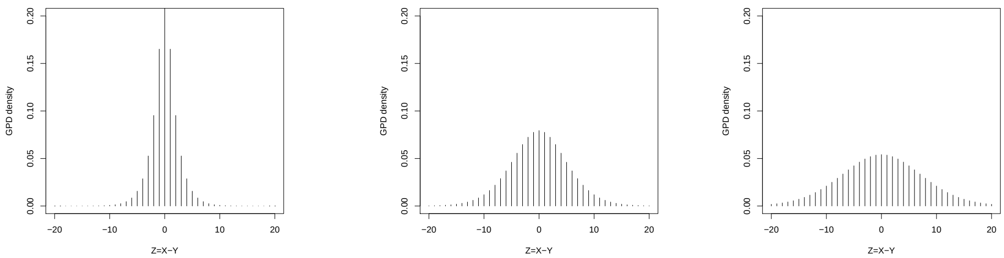
Figure 2.2: GPD density $\theta = 0.3, \lambda_1 = 1$ and $\lambda_2 = 2, 4, 6$.



(a) $\lambda_1 = 1, \lambda_1 = 2, \theta = 0.3$ (b) $\lambda_1 = 1, \lambda_2 = 2, \theta = 0.5$ (c) $\lambda_1 = 1, \lambda_2 = 2, \theta = 0.6$

Figure 2.3: GPD density $\lambda_1 = 1, \lambda_2 = 2$ and $\theta = 0.3, 0.5, 0.6$.

The QB distribution was introduced by Consul (1974) as an urn model. A new urn model and a general expression to QB distribution was presented by Consul and Mital (1975). Consul (1975) considered the characterization of Lagrangian Poisson and quasi-binomial distributions and the relation between them. Fazal (1976) developed an optimal asymptotic hypothesis test to the parameter $\theta = 0$. Korwar (1977) characterized the QB distribution in the context of the Rao-Rubin damage model. Consul (1990) discussed some properties and applications of the QB distribution, considered its moments and maximum likelihood estimation and applied it to several data sets.



(a) $\lambda_1 = \lambda_2 = 1, \theta = 0.3$ (b) $\lambda_1 = \lambda_2 = 5, \theta = 0.3,$ (c) $\lambda_1 = \lambda_2 = 10, \theta = 0.3$

Figure 2.4: GPD density $\theta = 0.3$ and $\lambda_1 = \lambda_2 = 1, 5, 10.$

According to Fazal (1976), a discrete random variable Y is said to follow the QB distribution if its pmf is given by

$$P(Y = y) = p\,q\binom{m}{y}(p + y\,\theta)^{y-1}[q + (m-y)\theta]^{m-y-1}/(1 + m\,\theta)^{m-1}, \quad y = 0, 1, 2, \ldots, m,$$

where $\binom{m}{y} = \frac{m!}{(m-y)!y!}$, $0 \leq p < 1, q = 1 - p, 0 \leq \theta < 1.$

Remark 2.2.3. Note that for $\theta = 0$, we obtain the binomial distribution with parameters m and p as a particular case of quasi-binomial distribution.

The mean and the first factorial moment of the QB distribution are given by

$$\mathrm{E(Y)} = mp \quad \text{and} \quad \mathrm{E}\left[\mathrm{Y(Y-1)}\right] = m(m-1)p - pq\sum_{j=0}^{m-2}\frac{m!\theta^j}{(m-j-2)!(1+m\theta)^{j+1}}.$$

Thus, the variance is given by

$$\mathrm{Var(Y)} = pq\left[m^2 - \sum_{j=0}^{m-2}\frac{m!\theta^j}{(m-j-2)!(1+m\theta)^{j+1}}\right].$$

Shenton (2006) gave an overview on the QB distribution and some properties were listed. The QB distribution converges in distribution to the GP distribution as m increases. The sum of quasi-binomial variates in general is not a quasi-binomial variate and properties of the binomial distribution, such as expressions of the partial sum of probabilities in integral form, are not readily available for the quasi-binomial distribution.

Alzaid and Al-Osh (1993) introduced the first-order autoregressive process with generalized Poisson marginal distribution (GPAR(1)) based on the quasi-binomial operator and obtained some properties like mean, variance, autocorrelation function, conditional variance and conditional expectation. This process can be viewed as an extension of Poisson INAR(1) process by Al-Osh and Alzaid (1987).

Alzaid and Al-Osh (1993) introduced the quasi-binomial operators as a sequence of random variables $\{S(n); n = 0, 1, 2, \ldots\}$ such that $S(n)$ has $\mathrm{QB}(p, \theta/\lambda, n)$ distribution for some p, θ and λ. Weiß (2008) called this operator as *"quasi-binomial thinning operator"*.

The quasi-binomial thinning operator has an interpretation in analogy with the binomial thinning operator. In the same way of the binomial thinning in the development of the Poisson INAR(1), the thinned variable $S(X)$ can be understood as the number of survivors from the population behind X (Weiß, 2008). Unlike the assumption in the Poisson INAR(1) process that given $X_{t-1} = x$, the number of retained elements to time t is a random variable with binomial distribution $B(\alpha, x)$, in the development of the GPAR(1) process it was assumed that given $X_{t-1} = x$ the number of retained elements to time t has a quasi-binomial distribution $QB(p, \theta, x)$. An element of the Poisson process at time $t-1$, X_{t-1}, independently of the other elements and the time of the process has a constant probability of being retained to time t, in the GPAR(1) process the quasi-binomial thinning operator consider that the probability of retaining an element is not constant but might depend on the time and the number of elements already retained (Alzaid and Al-Osh, 1993).

Based on the following proposition, Alzaid and Al-Osh (1993) defined the first-order autoregressive process with generalized Poisson marginal distribution.

Proposition 2.2.4. *(Alzaid and Al-Osh, 1993) If X is distributed according to the GP(λ, θ) distribution and the quasi-binomial thinning is performed independently of X, then $S(X)$ is distributed according to GP$(p\lambda, \theta)$.*

The GPAR(1) process $\{X_t\}$ is defined as

$$X_t = S_t(X_{t-1}) + \epsilon_t, \qquad t \in \mathbb{N},$$

where $\{\epsilon_t\}$ is an i.i.d. sequence of random variables with GP$(q\lambda, \theta)$ and $\{S_t(\cdot), t = 1, 2, \ldots\}$ is a sequence of i.i.d. quasi-binomial operators independent of ϵ_t with QB$(p, \theta/\lambda, \cdot)$ distribution, with $\lambda > 0,\ 0 < \theta < 1,\ 0 < p < 1,\ q = 1 - p$.

We have from the definition that $S_t(X_{t-1})|\,(X_{t-1} = x)$ has $\mathrm{QB}(p, \theta/\lambda, x)$ distribution. Moreover, assuming that X_0 follows a $\mathrm{GP}(\lambda, \theta)$ distribution independent of $\{\epsilon_t\}$ and $\{S_t(\cdot)\}$, the GPAR(1) process is a strictly stationary Markov chain with generalized Poisson marginal distribution (Alzaid and Al-Osh, 1993).

The mean and variance of the GPAR(1) process are given in (2.2), the coefficients of skewness and kurtosis are given in (2.3). It is not difficult to verify that the conditional mean and variance of the GPAR(1) process given X_{t-1} are, respectively,

$$\mathrm{E}(X_t|X_{t-1} = x) = px + \mu_\epsilon$$

and

$$\mathrm{Var}(X_t|X_{t-1}=x) = pq\left[x^2 - \sum_{j=1}^{x-2}\frac{x!\theta^j}{(x-j-2)!(1+x\theta)^{j+1}}\right] + \sigma_\epsilon^2,$$

where $\mu_\epsilon = \mathrm{E}(\epsilon_t) = q\lambda/(1-\theta)$ and $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_t) = q\lambda/(1-\theta)^3$.

The autocorrelation function of the GPAR(1) is given by

$$\rho_X(k) = \mathrm{corr}(X_t, X_{t-k}) = p^{|k|}, \quad k \in \mathbb{Z}.$$

Note that the autocorrelation function only depends on the p parameter and its behaviour is similar to that of the AR(1). Another characteristic of the GPAR(1) process is that its variance is greater than the mean while the Poisson INAR(1) process has the mean equal to the variance.

Al-Nachawati et al. (1997) compared two methods of estimating the parameters of the GPAR(1): Yule-Walker and Gaussian estimation. They observe that, for $\lambda > 1$, in small and larger sample sizes the Gaussian method has better results than YW estimator according the bias and MSE of estimates. For $\lambda \leq 1$, in small and larger sample sizes the YW and Gaussian methods presented similar results in terms of bias and MSE of estimates.

Now, based on the quasi-binomial operator, we will define the new operator $R(\cdot)$ and, in the same way to Freeland (2010), we define the new first order autoregressive process with generalized Poisson difference distribution marginals. The main feature of the GPDAR(1) process is more flexible in tails than the TINAR process (Freeland, 2010), because the GPD distribution may have longer or shorter tails (Shahtahmassebi and Moyeed, 2014). The process is skew or symmetric and has positive and negative values, the sequence of innovations has a generalized Poisson difference distribution, as the marginal of the process.

Definition 2.2.1. *Let X and Y be two independent random variables with $GP(\lambda_1,\theta)$ and $GP(\lambda_2,\theta)$ distributions, respectively, $Z = X - Y$ following GPD$(\lambda_1,\lambda_2,\theta)$ distribution. We define the operator $R(\cdot)$ as*

$$R(Z)|Z \stackrel{d}{=} S(X) - S(Y)|(X-Y), \tag{2.4}$$

where $S(X)|\,(X=x)$ and $S(Y)|\,(Y=y)$ have $\mathrm{QB}(p,\theta/\lambda_1,x)$ and $\mathrm{QB}(p,\theta/\lambda_2,y)$ distributions, respectively and "$\stackrel{d}{=}$" mean distribution equality.

From (2.4) we have that $R(Z) \stackrel{d}{=} S(X) - S(Y)$. Let X and Y be two independent random variables, so $S(X)$ and $S(Y)$ are independent random variables. From the Proposition 2.2.4 of Alzaid and Al-Osh (1993) of the quasi-binomial operator we have that $S(X)$ and $S(Y)$ have GP distributions.

Let $\mathbb{N}_0 = \{0,1,2,\ldots\}$. Let $\{X_t\}_{t\in\mathbb{N}_0}$ and $\{Y_t\}_{t\in\mathbb{N}_0}$ be two independent GPAR(1) processes with Generalized Poisson marginals, $\mathrm{GP}(\lambda_1,\theta)$ and $\mathrm{GP}(\lambda_2,\theta)$ distributions, respectively,

$$X_t = S_t(X_{t-1}) + \epsilon_t, t \in \mathbb{N} \tag{2.5}$$

and

$$Y_t = S_t(Y_{t-1}) + \delta_t, t \in \mathbb{N}. \tag{2.6}$$

Furthermore, the sequences $S_t(X_{t-1})|\,(X_{t-1} = x)$ and $S_t(Y_{t-1})|\,(Y_{t-1} = y)$ follow $\mathrm{QB}(p, \theta/\lambda_1, x)$, $\mathrm{QB}(p, \theta/\lambda_2, y)$ distributions independent of $\{\epsilon_t\}_{t\in\mathbb{N}}$ and $\{\delta_t\}_{t\in\mathbb{N}}$, with $\lambda_1 > 0$, $\lambda_2 > 0$, $0 < p < 1$, $0 < \theta < 1$, $q = 1 - p$. The sequences $\{\epsilon_t\}_{t\in\mathbb{N}}$ and $\{\delta_t\}_{t\in\mathbb{N}}$ are independent with $\mathrm{GP}(q\lambda_1, \theta)$ and $\mathrm{GP}(q\lambda_2, \theta)$ distributions, respectively.

The integer-valued first-order autoregressive process with generalized Poisson difference marginals is defined as follow.

Definition 2.2.2. *Let X_t and Y_t defined as above. We define the integer-valued first-order autoregressive process with generalized Poisson difference marginals, GPDAR(1), $\{Z_t\}_{t\in\mathbb{N}_0}$, as a sequence of a random variables such that*

$$Z_t \stackrel{d}{=} X_t - Y_t \stackrel{d}{=} R_t(Z_{t-1}) + \nu_t, \tag{2.7}$$

where $R_t(Z_{t-1})$ is a sequence of operators defined as (2.4) and $\nu_t = \epsilon_t - \delta_t$.

We have from (2.5) and (2.6) that ϵ_t and δ_t follow $\mathrm{GP}(q\lambda_1, \theta)$ and $\mathrm{GP}(q\lambda_2, \theta)$ distributions, respectively, so we have that ν_t follows the generalized Poisson difference distribution $\mathrm{GPD}(q\lambda_1, q\lambda_2, \theta)$.

Remark 2.2.5. This process can be viewed as an extension of the PDINAR(1) proposed by Alzaid and Omair (2014).

2.3 Properties of the process

In this section, we shall present some properties of the GPDAR(1) process, such as mean, variance, skewness and kurtosis coefficients, probability generating and cumulant generating functions. The autocorrelation and spectral density functions are also obtained. Follow, we present the marginal distribution properties of the GPDAR(1) process and the distribution properties of the sequence $\{\nu_t\}$ can be obtained in the same way of it.

The mean and variance of Z_t are given, respectively, by

$$\mu_Z = \mathrm{E}(Z_t) = \mathrm{E}(X_t) - \mathrm{E}(Y_t) = \frac{\lambda_1 - \lambda_2}{1 - \theta} \tag{2.8}$$

and

$$\sigma_Z^2 = \mathrm{Var}(Z_t) = \mathrm{Var}(X_t) + \mathrm{Var}(Y_t) = \frac{\lambda_1 + \lambda_2}{(1 - \theta)^3}. \tag{2.9}$$

Remark 2.3.1. Note that $|\mu_Z| < \sigma_Z^2$, then the GPDAR(1) process can be used as a model for overdispersed integer valued time series.

Consul (1986) considered the difference of two independent random variables with generalized Poisson distribution and obtained some properties. From those we have that the pgf of Z_t is given by

$$G_Z(s) = G_X(s) \cdot G_Y(s) = \mathrm{e}^{[\lambda_1(t_1-1)+\lambda_2(t_2-1)]},$$

where $G_X(s)$ and $G_Y(s)$ are the X_t and $-Y_t$ pgf's and $t_1 = s\,\mathrm{e}^{\theta(t_1-1)}$, $t_2 = s^{-1}\mathrm{e}^{\theta(t_2-1)}$.

The cumulant generating function (cgf) of Z_t is given by

$$\psi(\beta) = \lambda_1(\mathrm{e}^{s_1} - 1) + \lambda_2(\mathrm{e}^{s_2} - 1) = (s_1 - \beta)\lambda_1/\theta + (s_2 + \beta)\lambda_2/\theta,$$

where $s_1 = \beta + \theta(\mathrm{e}^{s_1} - 1)$, $s_2 = -\beta + \theta(\mathrm{e}^{s_2} - 1)$, and the first four cumulants are

$$k_1 = \frac{\lambda_1 - \lambda_2}{1-\theta}, \qquad k_2 = \frac{\lambda_1 + \lambda_2}{(1-\theta)^3},$$
$$k_3 = \frac{(\lambda_1 - \lambda_2)(1+2\theta)}{(1-\theta)^5}, \qquad k_4 = \frac{(\lambda_1 + \lambda_2)(1+8\theta+6\theta^2)}{(1-\theta)^7}.$$

The coefficients of skewness and kurtosis of Z_t are given by

$$\gamma_1 = \frac{(\lambda_1 - \lambda_2)(1+2\theta)}{\sqrt{(\lambda_1+\lambda_2)^3}\sqrt{(1-\theta)}} \quad \text{and} \quad \gamma_2 = 3 + \frac{1+8\theta+6\theta^2}{(\lambda_1+\lambda_2)(1-\theta)},$$

respectively.

Remark 2.3.2. Note that if $\lambda_1 = \lambda_2$, we have a symmetric process with zero mean.

It is possible to show that the autocovariance function of Z_t process is given by

$$\gamma(h) = \operatorname{cov}(Z_t, Z_{t-h}) = \frac{\lambda_1 + \lambda_2}{(1-\theta)^3} p^{|h|},\ h \in \mathbb{Z},$$

and the autocorrelation function (ACF) at lag h is given by

$$\rho(h) = \operatorname{Corr}(Z_t, Z_{t-h}) = p^{|h|}, \quad h \in \mathbb{Z}.$$

Since $0 < p < 1$, the autocorrelation function decays exponentially, like in the AR(1) process.

Remark 2.3.3. The process defined in (2.7) has positive autocorrelation. We can define a GPDAR(1) process with negative autocorrelation. In the same way of Freeland (2010), we define

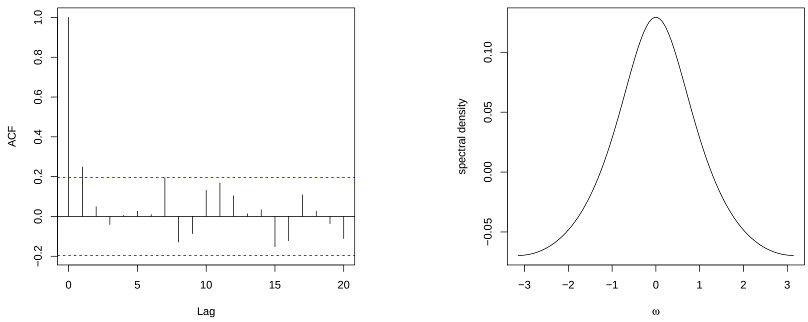
$$Z_t = \begin{cases} X_t - Y_t, & t = 0, 2, 4, \ldots \\ Y_t - X_t, & t = 1, 3, 5, \ldots \end{cases}$$

Then, we have in this case that $\rho(h) = \mathrm{Corr}(Z_t, Z_{t-h}) = (-p)^h$, for $h \in \mathbb{N}$. The main properties of this process can be obtained as in the process defined in (2.7).

The spectral density function $f(\omega)$ of the GPDAR(1) process is

$$f(\omega) = \frac{(\lambda_1 + \lambda_2)}{2\pi(1-\theta)^3} \frac{2p(\cos(\omega) - p)}{(1 + p^2 - 2p\cos(\omega))}, \quad \omega \in (-\pi, \pi].$$

Figure 2.5 displays a sample autocorrelation based on 100 simulated values and the spectral density function for $p = 0.3$, $\lambda_1 = 1$, $\lambda_2 = 2$ and $\theta = 0.8$.



(a) Sample autocorrelation (b) Spectral density

Figure 2.5: Sample autocorrelation based on 100 simulated values of the GPDAR(1) and spectral density function for $p = 0.3$, $\lambda_1 = 1$, $\lambda_2 = 2$ and $\theta = 0.8$.

Now, we present two results about the r.v. $R_t(Z_{t-1})$ and the GPDAR(1) process. In Proposition 2.3.4, we present the conditional probability function of the r.v. $R_t(Z_{t-1})$ given Z_{t-1}, this result is particularly useful if our interest is to obtain the transition probabilities from the process. In Proposition 2.3.5, the conditional expectation and variance of GPDAR(1) process are obtained. The conditional expectation expression is directly linked to 1-step ahead forecasts and used the conditional least squares estimators for unknown parameters of the process. The proof of the propositions can be found in the Appendix.

Proposition 2.3.4. The conditional probability function of the random variable $R_t(Z_{t-1})$ given Z_{t-1} is given by

$$P(R_t(Z_{t-1}) = j | Z_{t-1} = k) = \sum_{l=0}^{\infty} \sum_{i=0}^{\min(l, l+k-j)} \frac{(\lambda_1,\theta)_{k+l} (\lambda_2,\theta)_l \, \mathrm{e}^{-2l\theta} f(i+j; p, \theta/\lambda_1, k+l) f(i; p, \theta/\lambda_2, l)}{\sum_{y=0}^{\infty} (\lambda_2,\theta)_y (\lambda_1,\theta)_{y+k} \mathrm{e}^{-2y\theta}},$$

for $j \geq 0, k > 0$, and

$$P(R_t(Z_{t-1}) = j|Z_{t-1} = k) = \sum_{l=0}^{\infty} \sum_{i=0}^{\min(l-j,l-k)} \frac{(\lambda_1,\theta)_l(\lambda_2,\theta)_{l-k}\, \mathrm{e}^{-2l\theta} f(i;p,\theta/\lambda_1,l) f(i+j;p,\theta/\lambda_2,l-k)}{\sum_{y=0}^{\infty}(\lambda_2,\theta)_y(\lambda_1,\theta)_{y+k}\mathrm{e}^{-2(y+k)\theta}},$$

for $j \geq 0, k < 0$, where $(\lambda, \theta)_y = \dfrac{\lambda(\lambda + y\theta)^{y-1}}{y!}$ and $f(i; p, \theta, n)$ is the QB's probability mass function with parameters (p, θ, n).

Proposition 2.3.5. Let $\{Z_t\}_{t\in\mathbb{N}_0}$ be a GPDAR(1) process according to Definition 2.2.2. The conditional expectation of Z_t given Z_{t-1} is given by

$$\mathrm{E}(Z_t|Z_{t-1}) = pZ_{t-1} + \mu_\nu,$$

where $\mu_\nu = \mathrm{E}(\nu_t) = \mathrm{E}(\epsilon_t) - \mathrm{E}(\delta_t) = \dfrac{q(\lambda_1 - \lambda_2)}{1 - \theta}$ and the conditional variance is given by

$$\mathrm{Var}(Z_t|Z_{t-1} = z) = \sum_{y=0}^{\infty} \frac{(\lambda_2,\theta)_y(\lambda_1,\theta)_{y+z}\{p[(z+y)(z+y(1-p))+y^2-q[W_{(z+y)}+W_{(y)}]\}\mathrm{e}^{-2y\theta}}{\sum_{y=0}^{\infty}(\lambda_2,\theta)_y(\lambda_1,\theta)_{z+y}\mathrm{e}^{-2y\theta}} - (pz + \mu_\nu)^2,$$

for $z \geq 0$, and

$$\mathrm{Var}(Z_t|Z_{t-1} = z) = \sum_{x=0}^{\infty} \frac{(\lambda_2,\theta)_{(x-z)}(\lambda_1,\theta)_x\{p[(x-z)(x(1-p)-z)+x^2-q(W_{(x)}+W_{(x-z)})]\}\mathrm{e}^{-2x\theta}}{\sum_{y=0}^{\infty}(\lambda_2,\theta)_y(\lambda_1,\theta)_{y+z}\mathrm{e}^{-2(y+z)\theta}} - (pz+\mu_\nu)^2,$$

for $z < 0$. Where $W_{(x)} = \sum_{j=0}^{x-1} \dfrac{x!\theta^j}{(x-j-2)!(1+x\theta)^{j+1}}$ and $(\lambda, \theta)_y$ was defined as in Proposition 2.3.4.

Remark 2.3.6. The expressions of the conditional variance are not analytically simple, but it is not difficult to evaluate them numerically.

2.4 Estimation

In this section, we outline the estimation of the unknown parameters of the GPDAR(1) process. We investigate the estimation method of moments with Yule-Walker .

Considering the skew process, $\lambda_1 \neq \lambda_2$, we can use Yule-Walker estimation to estimate p. From a T-points observation $Z_1, Z_2, \ldots, Z_T$, based on the fact that the correlation function is given by $\rho(h) = p^{|h|}$ we consider the Yule-Walker estimator given by

$$\widehat{p} = \widehat{\rho}(1) = \frac{\sum_{t=1}^{T-1}(Z_t - \bar{Z})(Z_{t+1} - \bar{Z})}{\sum_{t=1}^{T}(Z_t - \bar{Z})^2}.$$

Based on the method of moments, using the sample mean and variance $\bar{Z} = \frac{1}{T}\sum_{t=1}^{T} Z_t$, $\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(Z_t - \bar{Z})^2$, and the two first moments given by (2.8) and (2.9), we have

$$\bar{Z} = \frac{\widehat{\lambda}_1 - \widehat{\lambda}_2}{1 - \widehat{\theta}} \text{ and } \widehat{\sigma}^2 = \frac{\widehat{\lambda}_1 + \widehat{\lambda}_2}{(1 - \widehat{\theta})^3}.$$

Then, after a simple algebra, the estimators for λ_1 and λ_2 are given by

$$\widehat{\lambda}_1 = \frac{1}{2}\left[\widehat{\sigma}^2(1 - \widehat{\theta})^3 + \bar{Z}(1 - \widehat{\theta})\right]$$

and

$$\widehat{\lambda}_2 = \frac{1}{2}\left[\widehat{\sigma}^2(1 - \widehat{\theta})^3 - \bar{Z}(1 - \widehat{\theta})\right].$$

Using the sample third moment $m_3 = \frac{1}{T}\sum_{t=1}^{T}(Z_t - \bar{Z})^3$ and the third cumulant, the moment estimator of θ satisfies the equation

$$(1 - \widehat{\theta})^4 m_3 - \bar{Z}(1 + 2\widehat{\theta}) = 0. \tag{2.10}$$

It is not difficult to show that, if m_3 and $\overline{Z}$ are both positive or negative, Equation (2.10) has a solution in $(0, 1)$ and we can find this solution numerically using some computer software, for example, the R software (see http://www.r-project.org). Note that the moment estimators for λ_1 and λ_2 are well defined if $\widehat{\sigma}^2(1 - \widehat{\theta})^2 >\mid \overline{Z} \mid$. Furthermore, if $\lambda_1 = \lambda_2$, we have a symmetric process with zero mean, so the estimators presented above do not have sense in this case and is necessary an estimator more adequate.

Consider the symmetric process with $\lambda_1 = \lambda_2 = \lambda$. Using the sample variance $\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(Z_t - \overline{Z})^2$ and the variance of the process, $\sigma_Z^2 = 2\lambda/(1 - \theta)^3$, the moment estimator for λ is given by

$$\widehat{\lambda} = \frac{1}{2}\widehat{\sigma}^2(1 - \widehat{\theta})^3.$$

Now, consider the fourth cumulant k_4 and variance σ^2, defined in Section 2.3, and the well known relation $k_4 = \mu_4 - 3\mu_2^2$, where μ_4 and μ_2 are the fourth and second central moment. Using the second and fourth sample central moments, $m_2 = \frac{1}{T}\sum_{t=1}^{T}(Z_t - \overline{Z})^2$ and $m_4 = \frac{1}{T}\sum_{t=1}^{T}(Z_t - \overline{Z})^4$, respectively, the moment estimator of θ satisfies

$$(1 - \widehat{\theta})^4 \widehat{k}_4 - \widehat{\sigma}^2(1 + 8\,\widehat{\theta} + 6\,\widehat{\theta}^2) = 0, \tag{2.11}$$

where $\widehat{k}_4 = m_4 - 3m_2^2$. The equation (2.11) has solution since that $\widehat{k}_4 > \widehat{\sigma}^2$. The moment estimator of θ do not have closed form but we can find it numerically as in the last case.

The conditional least squares method consists of minimizing the function

$$Q_T(\boldsymbol{\beta}) = \sum_{t=2}^{T}[Z_t - E(Z_t|Z_{t-1})]^2,$$

where $\boldsymbol{\beta}$ is a vector of unknown parameters of the process. Here, the function to be minimized is given by

$$Q_T(p, \mu_\nu) = \sum_{t=2}^{T}(Z_t - pZ_{t-1} - \mu_\nu)^2 = \sum_{t=2}^{T}[Z_t - pZ_{t-1} - (1-p)\mu]^2.$$

Note that using this method we can not obtain estimators for λ_1, λ_2 and θ, but only p and $\mu = (\lambda_1 - \lambda_2)/(1-\theta)$, for this reason, this method of estimation was not considered.

2.5 Simulation study

A Monte Carlo simulation experiment was performed to study the behaviour of the Yule-Walker estimators. The number of Monte Carlo replications was 5000 and the considered sample sizes were T = 50, 100 and 200. For the values of parameters, we considered $p = 0.5, 0.6, \lambda_1 = 1, 2, \lambda_2 = 1, 2$ and $\theta = 0.6$. The Monte Carlo simulation was performed using the R programming language, see http://www.r-project.org.Table 1 presents the bias and mean squared error (MSE) of the estimates of the parameters of the GPDAR(1) process. From the results we conclude that estimates of the parameters are reasonable, there are convergence to the true parameters values, and the bias and MSE decrease as the size sample increases as expected. The bias and MSE of λ_1 and λ_2 estimates decreases non-uniformly. Note that the MSE values of λ_1 and λ_2 estimators are large for $\theta = 0.6$, this is justified because the random sample will have a large variance and this estimators depends on it directly.

2.6 Application to real data set

In this section, we present an application of the GPDAR(1) model to the GLOBAL Land-Ocean temperature index. We consider the monthly GLOBAL Land-Ocean temperature index in 0.01 degrees Celsius for 1950-1962 update december 2015. The temperature index in 0.01 degree Celsius belongs to $\mathbb{Z}$ and is our time series of interest here in this section. The data set consist of 156 observations. The data has been downloaded from the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce, Earth System Research Laboratory (`http://www.esrl.noaa.gov/psd/data/climateindices/list`), originally from `http://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts.txt`.

Table 2.2 provides some descriptive measures for the temperature index, which include central tendency statistics, variance, skewness and kurtosis, among others. We see that the data set assumes positive and negative integer values.

Table 2.1: Bias and mean squared error (in parentheses) of the moment estimates of the parameters for $p = 0.5, 0.6$ $\lambda_1 = 1, 2$, $\lambda_2 = 1, 2$, $\theta = 0.6$ and $T = 50, 100, 200$.

T	$\widehat{p}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\theta}$
	$p = 0.6, \lambda_1 = 1.0, \lambda_2 = 2.0$ and $\theta = 0.6$			
50	−0.0960 (0.0362)	0.1257 (2.6563)	0.3166 (4.2966)	−0.0211 (0.0188)
100	−0.0543 (0.0195)	0.0946 (2.2423)	0.2017 (3.0628)	−0.0032 (0.0130)
200	−0.0329 (0.0099)	0.1626 (2.9754)	0.2134 (3.6921)	0.0024 (0.0109)
	$p = 0.5, \lambda_1 = 1.0, \lambda_2 = 2.0$ and $\theta = 0.6$			
50	−0.0842 (0.0365)	0.1204 (2.6663)	0.2988 (4.1294)	−0.0138 (0.0168)
100	−0.0504 (0.0198)	0.1396 (2.6824)	0.2313 (3.5838)	−0.0040 (0.0125)
200	−0.0295 (0.0109)	0.1599 (2.7864)	0.1997 (3.3934)	0.0033 (0.0101)
	$p = 0.6, \lambda_1 = 2.0, \lambda_2 = 1.0$ and $\theta = 0.6$			
50	−0.0971 (0.0371)	0.3248 (5.2681)	0.1466 (3.5077)	−0.0223 (0.0192)
100	−0.0579 (0.0197)	0.1889 (2.9672)	0.0841 (2.0812)	−0.0038 (0.0140)
200	−0.0328 (0.0104)	0.2081 (3.5633)	0.1566 (2.8898)	0.0038 (0.0106)
	$p = 0.5, \lambda_1 = 2.0, \lambda_2 = 1.0$ and $\theta = 0.6$			
50	−0.0908 (0.0381)	0.2167 (2.8068)	0.0594 (1.6708)	−0.0142 (0.0168)
100	−0.0500 (0.0199)	0.2812 (3.9913)	0.1786 (3.0712)	−0.0053 (0.0133)
200	−0.0287 (0.0104)	0.1999 (3.3034)	0.1577 (2.7065)	0.0041 (0.0096)

Table 2.2: Descriptive statistics.

Minimum	Median	Mean	Variance	Skewness	Kurtosis	$\widehat{\rho}(1)$	Maximum
-47	-2.50	-4.07	260.8	-0.245	2.60	0.523	38

The time series data and their sample autocorrelations and partial autocorrelations are displayed in Figure 2.6. Analyzing Figure 2.6, we conclude that a first order autoregressive model may be appropriate for the given data series, because of the clear cut-off after lag 1 in the partial autocorrelations. Furthermore, the behavior of the series indicates that it may be mean stationary. We compare our model with the TINAR(1) process introduced by Freeland (2010) (asymmetric version).

The TINAR(1) process considered here has marginals following a Skellam distribution with parameters $\lambda_1(1-p)^{-1}$ and $\lambda_2(1-p)^{-1}$, i.e., the marginals are distributed as $Y_1 - Y_2$, where Y_1 and Y_2 are two independent Poisson random variables with mean $\lambda_1(1-p)^{-1}$ and $\lambda_2(1-p)^{-1}$, respectively, and p is the associated thinning parameter of this process.

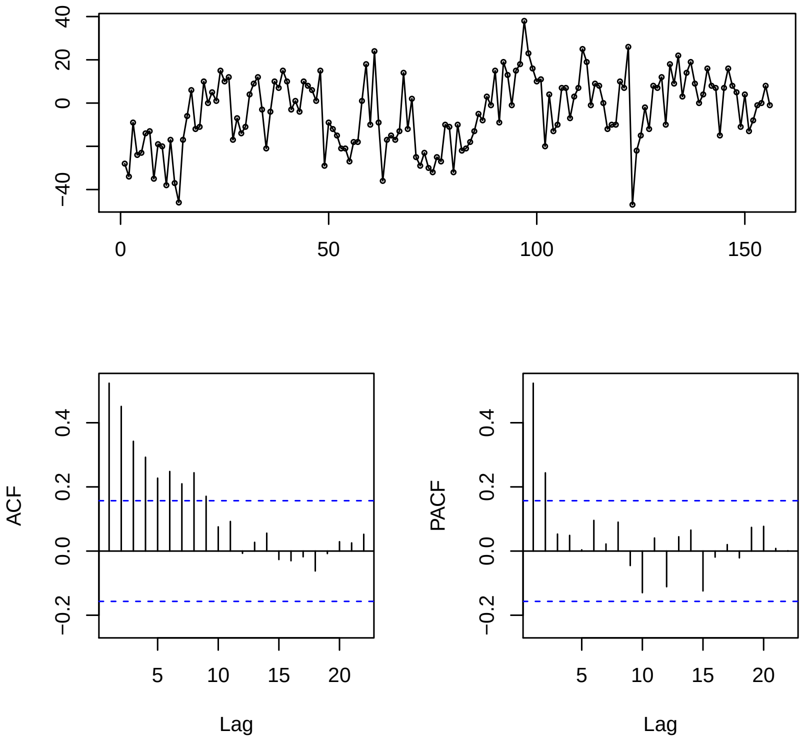


Figure 2.6: Time series plot, autocorrelation and partial autocorrelation functions for the GLOBAL Land-Ocean temperature index.

In Table 4.6 we present the estimates of the parameters with their MdAE (median absolute error) and estimated quantities (index of dispersion, skewness and kurtosis) for the fitted models. The MdAE statistic is defined as follows. For $t = 2, \ldots, T$, consider the expected value of the observation at the previous time, $\mathrm{E}[Z_t|Z_{t-1}] = p\, Z_{t-1} + (1-p)(\lambda_1 - \lambda_2)/(1-\theta)$ (for GPDAR(1) process). The MdA is obtained as the median of the $|Z_t - \mathrm{E}(Z_t|Z_{t-1})|$. In general it is expected that the best model to fit data presents the smallest values for this statistic. From Table 4.6, note that all two models exhibit good fit of mean and variance, but only the GPDAR(1) model give a reasonable fit of skewness and kurtosis, i.e., the TINAR(1) model is not able to reproduce the skewness and kurtosis. Furthermore, the values of MdAE statistic indicate that the GPDAR(1) model showed better result compared with TINAR(1) model. Based on these facts, we conclude that the GPDAR(1) model capture more information of these data. The sample autocorrelations of the residuals obtained from GPDAR(1) model are shown in Figure 2.7. From this figure, we conclude that the proposed model fitted the data well.

Table 2.3: Estimates of the parameters, MdAE and estimated quantities for the GPDAR(1) and TINAR(1) processes.

Model	Estimates	MdAE	σ_Z^2/μ_Z	γ_1	γ_2
GPDAR(1)	$\widehat{p} = 0.5233$	8.6871	-64.071	-0.2433	3.7966
	$\widehat{\lambda}_1 = 3.3152$				
	$\widehat{\lambda}_2 = 4.5841$				
	$\widehat{\theta} = 0.6882$				
TINAR(1)	$\widehat{p} = 0.5233$	8.6909	-64.071	-0.0010	0.0038
	$\widehat{\lambda}_1 = 61.1691$				
	$\widehat{\lambda}_2 = 63.1088$				
Empirical			-64.071	-0.2433	2.5993

2.7 Concluding remarks

In this chapter, we introduced the GPDAR(1) process. The main advantages of this new model is the skewness, possible negative values for the time series and the distribution of the difference of two GP random variables can have longer or shorter tails compared to the PD distribution and the GPD distribution can be viewed as a generalisation of the PD distribution. We did a review of the GP, quasi-binomial distribution and the GPAR(1) process. The properties of the GPDAR(1) process was presented, mean, variance, skewness and kurtosis coefficients, probability generating and cumulant function. The autocorrelation and spectral density are

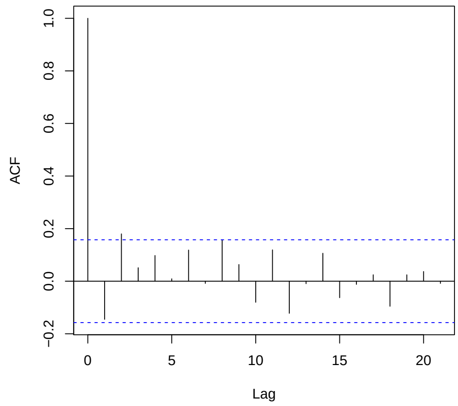


Figure 2.7: Sample autocorrelations of the residuals obtained from GPDAR(1) model.

obtained. The conditional expectation and variance were derived. A Monte Carlo simulation study shows that the Yule-Walker estimators have a reasonable behaviour. An application to real data set shows that the GPDAR(1) process is adequate to fit skew data set with positive and negative values.

2.8 Appendix

Proof of Proposition 2.3.4

First, let us consider the case $j \geq 0$ and $k \geq 0$. We have,

$$P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j | Z_{t-1} = k) = \frac{P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j, Z_{t-1} = k)}{P(Z_{t-1} = k)}$$

$$= \frac{1}{P(Z_{t-1}=k)} \sum_{l=0}^{\infty} P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j | X_{t-1} = k+l, Y_{t-1} = l) P(X_{t-1} = k+l, Y_{t-1} = l).$$

We know that the random variables X_{t-1} and Y_{t-1} have $GP(\lambda_1, \theta)$ and $GP(\lambda_2, \theta)$ distributions, respectively. So, we find $P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j | X_{t-1} = k+l, Y_{t-1} = l)$. Define $U = S_t(X_{t-1}) | X_{t-1} = k+l$ and $V = S_t(Y_{t-1}) | Y_{t-1} = l$, with $QB(p, \theta/\lambda_1, k+l)$ and $QB(p, \theta/\lambda_2, l)$ distributions respectively. X_t and Y_t are independent random variables, so the conditional distribution $P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j | X_{t-1} = k+l, Y_{t-1} = l)$ is the convolution of two random variables with QB distributions.

Thus, let $*$ denote convolution, let $f(i;p,\theta,n)$ be the QB's pmf, this is

$$f(i;p,\theta,n) = \frac{pq\binom{n}{i}(p+y\theta)^{y-1}[q+(n-y)\theta]^{n-y-1}}{(1+n\theta)^{n-1}}.$$

Then,

$$P(U-V=j) = \sum_i f(i+j;p,\theta/\lambda_1,k+l)f(i;p,\theta/\lambda_2,l).$$

We have

$$0 \le i \le l \text{ and } i \le k+l-j \Rightarrow 0 \le i \le \min(l,l+k-j).$$

Thus,

$$P(U-V=j) = \sum_{i=0}^{\min(l,l+k-j)} f(i+j;p,\theta/\lambda_1,k+l)f(i;p,\theta/\lambda_2,l)$$

and

$P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j|Z_{t-1} = k)$

$= \frac{1}{P(Z_{t-1}=k)} \sum_{l=0}^{\infty} P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j|X_{t-1} = k+l, Y_{t-1} = l)P(X_{t-1} = k+l, Y_{t-1} = l)$

$= \sum_{l=0}^{\infty} \frac{P(X_{t-1}=k+l,Y_{t-1}=l)}{P(Z_{t-1}=k)} \sum_{i=0}^{\min(l,l+k-j)} f(i+j;p,\theta/\lambda_1,k+l)f(i;p,\theta/\lambda_2,l)$

$= \sum_{l=0}^{\infty} \sum_{i=0}^{\min(l,l+k-j)} \frac{(\lambda_1,\theta)_{k+l}(\lambda_2,\theta)_l \, \mathrm{e}^{-2l\theta} f(i+j;p,\theta/\lambda_1,k+l)f(i;p,\theta/\lambda_2,l)}{\sum_{y=0}^{\infty}(\lambda_2,\theta)_y(\lambda_1,\theta)_{y+k}\mathrm{e}^{-2y\theta}}.$

For $j \ge 0, k < 0$ we have

$P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j|Z_{t-1} = k) = \frac{P(S_t(X_{t-1})-S_t(Y_{t-1})=j,Z_{t-1}=k)}{P(Z_{t-1}=k)}$

$= \frac{1}{P(Z_{t-1}=k)} \sum_{l=0}^{\infty} P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j|X_{t-1} = l, Y_{t-1} = l-k)P(X_{t-1} = l, Y_{t-1} = l-k).$

In this case, we have

$$i \le l-j \text{ and } 0 \le i \le l-k \Rightarrow 0 \le i \le \min(l-j,l-k)$$

and $P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j|Z_{t-1} = k)$

$= \dfrac{1}{P(Z_{t-1}=k)} \sum_{l=0}^{\infty} P(S_t(X_{t-1}) - S_t(Y_{t-1}) = j | X_{t-1} = l, Y_{t-1} = l-k) P(X_{t-1} = l, Y_{t-1} = l-k)$

$= \sum_{l=0}^{\infty} \dfrac{P(X_{t-1} = l, Y_{t-1} = l-k)}{P(Z_{t-1}=k)} \sum_{i=0}^{\min(l-j,l-k)} f(i+j; p, \theta/\lambda_1, l) f(i; p, \theta/\lambda_2, l-k)$

$= \dfrac{\sum_{l=0}^{\infty} \sum_{i=0}^{\min(l-j,l-k)} (\lambda_1,\theta)_l (\lambda_2,\theta)_{l-k}\, e^{-2l\theta} f(i+j; p, \theta/\lambda_1, l) f(i; p, \theta/\lambda_2, l-k)}{\sum_{y=0}^{\infty} (\lambda_2,\theta)_y (\lambda_1,\theta)_{y+k} e^{-2(y+k)\theta}}.$

Proof of Proposition 2.3.5

Using the definition of Z_t, we have

$$\begin{aligned} E(Z_t|Z_{t-1}=z) &= E(S_t(X_{t-1}) - S_t(Y_{t-1}) + \nu_t | X_{t-1} - Y_{t-1} = z) \\ &= E(S_t(X_{t-1}) - S_t(Y_{t-1}) | X_{t-1} - Y_{t-1} = z) + E(\nu_t) \end{aligned}$$

since that ν_t, X_{t-1} and Y_{t-1} are independent random variables and

$$E(\nu_t) = E(\epsilon_t) - E(\delta_t) = \frac{q\lambda_1}{1-\theta} - \frac{q\lambda_2}{1-\theta} = \frac{q(\lambda_1 - \lambda_2)}{1-\theta},$$

to find the conditional expectation of the Z_t given $Z_{t-1} = z$ we have find

$$E(S_t(X_{t-1}) + S_t(Y_{t-1}) | X_{t-1} - Y_{t-1} = z).$$

For $z \geq 0$ we have,

$$\begin{aligned} & E(S_t(X_{t-1}) - S_t(Y_{t-1}) | X_{t-1} - Y_{t-1} = z) \\ &= \sum_{r=-\infty}^{\infty} r P(S_t(X_{t-1}) - S_t(Y_{t-1}) | X_{t-1} - Y_{t-1} = z) \\ &= \sum_{r=-\infty}^{\infty} \sum_{y=0}^{\infty} r P(S_t(X_{t-1}) - S_t(Y_{t-1}) | X_{t-1} = z+y, Y_{t-1} = y) \\ &= \sum_{r=-\infty}^{\infty} \sum_{y=0}^{\infty} r \frac{P(S_t(X_{t-1}) - S_t(Y_{t-1}), X_{t-1} = z+y, Y_{t-1} = y)}{P(X_{t-1} - Y_{t-1} = z)} \\ &= \frac{\sum_{y=0}^{\infty} P(X_{t-1} = z+y) P(Y_{t-1} = y)}{P(X_{t-1} - Y_{t-1} = z)} \sum_{r=-\infty}^{\infty} r P(S_t(X_{t-1}) - S_t(Y_{t-1}) | X_{t-1} = z+y, Y_{t-1} = y) \\ &= \frac{\sum_{y=0}^{\infty} P(X_{t-1} = z+y) P(Y_{t-1} = y)}{P(X_{t-1} - Y_{t-1} = z)} E(S_t(X_{t-1}) - S_t(Y_{t-1}) | X_{t-1} = z+y, Y_{t-1} = y). \end{aligned}$$

As defined in 2.2, we know that the random variables $S_t(X_{t-1})|X_{t-1} = z{+}y$ and $S_t(Y_{t-1})|Y_{t-1} = y$ have $QB(p, \theta/\lambda_1, z+y)$ and $QB(p, \theta/\lambda_2, y)$ distributions, respectively. Then,

$$E(S_t(X_{t-1}) - S_t(Y_{t-1})|X_{t-1} = z+y, Y_{t-1} = y) = (z+y)p - y\,p = z\,p$$

Then,

$$E(S_t(X_{t-1}) - S_t(Y_{t-1})|X_{t-1} - Y_{t-1} = z) = \frac{\sum_{y=0}^{\infty} P(X_{t-1} = z+y)P(Y_{t-1} = y)}{P(X_{t-1} - Y_{t-1} = z)}\, zp = zp.$$

For $z < 0$, we have

$$\begin{aligned}
&E(S_t(X_{t-1}) + S_t(Y_{t-1})|X_{t-1} - Y_{t-1} = z)\\
&= \sum_{r=-\infty}^{\infty} rP(S_t(X_{t-1}) - S_t(Y_{t-1})|X_{t-1} - Y_{t-1} = z)\\
&= \sum_{r=-\infty}^{\infty}\sum_{x=0}^{\infty} rP(S_t(X_{t-1}) - S_t(Y_{t-1})|X_{t-1} = x, Y_{t-1} = x - z)\\
&= \sum_{r=-\infty}^{\infty}\sum_{x=0}^{\infty} r\frac{P(S_t(X_{t-1}) - S_t(Y_{t-1}), X_{t-1} = x, Y_{t-1} = x-z)}{P(X_{t-1} - Y_{t-1} = z)}\\
&= \frac{\sum_{x=0}^{\infty} P(X_{t-1} = z+y)P(Y_{t-1} = y)}{P(X_{t-1} - Y_{t-1} = z)} \sum_{r=-\infty}^{\infty} rP(S_t(X_{t-1}) - S_t(Y_{t-1})|X_{t-1} = x, Y_{t-1} = x - z)\\
&= \frac{\sum_{x=0}^{\infty} P(X_{t-1} = x)P(Y_{t-1} = x-z)}{P(X_{t-1} - Y_{t-1} = z)} E(S_t(X_{t-1}) - S_t(Y_{t-1})|X_{t-1} = x, Y_{t-1} = x - z),
\end{aligned}$$

so, as above,

$$E(S_t(X_{t-1}) - S_t(Y_{t-1})|X_{t-1} = x, Y_{t-1} = x - z) = xp - (x-z)p = z\,p.$$

Then,

$$E(S_t(X_{t-1}) - S_t(Y_{t-1})|X_{t-1} - Y_{t-1} = z) = zp\,\frac{\sum_{x=0}^{\infty} P(X_{t-1} = x)P(Y_{t-1} = x-z)}{P(X_{t-1} - Y_{t-1} = z)} = z\,p.$$

Then, for $Z_{t-1} = z, z \in \mathbb{Z}$, we have

$$E(Z_t|Z_{t-1}) = pZ_{t-1} + \mu_\nu,$$

where $\mu_\nu = E(\nu_t)$.

To find conditional variance, $\mathrm{Var}(Z_t|Z_{t-1} = z) = E(Z_t^2|Z_{t-1} = z) - [E(Z_t|Z_{t-1} = z)]^2$, we need to find $E(Z_t^2|Z_{t-1} = z)$. In the same way of conditional expectation, for $z \geq 0$ we have

$$\begin{aligned}
&E(Z_t^2|Z_{t-1} = z) = E[(S_t(X_{t-1}) - S_t(Y_{t-1}))^2|X_{t-1} - Y_{t-1} = z]\\
&= \sum_{y=0}^{\infty} \frac{P(X_{t-1} = z+y)P(Y_{t-1} = y)}{P(X_{t-1} - Y_{t-1} = z)} E[((S_t(X_{t-1}) - S_t(Y_{t-1}))^2|X_{t-1} = z+y, Y_{t-1} = y)],
\end{aligned}$$

where,

$$E[(S_t(X_{t-1}) - S_t(Y_{t-1})^2|X_{t-1} = z+y, Y_{t-1} = y)]$$
$$= E(S_t^2(X_{t-1})|X_{t-1} = z+y, Y_{t-1} = y)) + E(S_t^2(Y_{t-1})|X_{t-1} = z+y, Y_{t-1} = y))$$
$$-E[S_t(X_{t-1})|X_{t-1} = z+y, Y_{t-1} = y)]E[S_t(Y_{t-1})|X_{t-1} = z+y, Y_{t-1} = y)]. \quad (2.12)$$

We know that if the random variable U has quasi-binomial distribution then Var(U) is given by (2.2), so

$$E(S_t^2(X_{t-1})|X_{t-1} = z+y) = \text{Var}(S_t(X_{t-1})|X_{t-1}) + E^2(S_t(X_{t-1})|X_{t-1})$$

$$= p\,q\left[(y+z)^2 - \sum_{j=0}^{z+y-1} \frac{(z+y)!\theta^j}{(z+y-j-2)!(1+(z+y)\theta)^{j+1}}\right] + (z+y)^2p^2$$

$$= p\,q(y+z)^2 - p\,qW_{(z+y)} + p^2(z+y)^2 \;=\; p[(z+y)^2 - qW_{(z+y)}] \quad (2.13)$$

and

$$E(S_t^2(Y_{t-1})|Y_{t-1} = y) = p(y^2 - qW_y), \quad (2.14)$$

but $W_{(x)} = \sum_{j=0}^{x-1} \frac{x!\theta^j}{(x-j-2)!(1+x\theta)^{j+1}}$. X_{t-1} and Y_{t-1} are independent random variables, therefore

$$E[S_t(X_{t-1})|X_{t-1} = z+y, Y_{t-1} = y)]E[S_t(Y_{t-1})|X_{t-1} = z+y, Y_{t-1} = y)]$$
$$= [p(z+y)][py] = y(z+y)p^2. \quad (2.15)$$

Replace (2.13), (2.14) and (2.15) in (2.12). We have

$$E(Z_t^2|Z_{t-1} = z)$$
$$= \sum_{y=0}^{\infty} \frac{P(X_{t-1} = z+y)P(Y_{t-1} = y)}{P(X_{t-1} - Y_{t-1} = z)} \{p[(z+y)^2 - qW_{(z+y)}] + p(y^2 - qW_{(y)}) - y(z+y)p^2\}$$
$$= \sum_{y=0}^{\infty} \frac{P(X_{t-1} = z+y)P(Y_{t-1} = y)}{P(X_{t-1} - Y_{t-1} = z)} \{p[(z+y)(z+y(1-p)) + y^2] - pq[W_{(z+y)} - W_{(y)}]\}.$$

Also,

$$\frac{P(X_{t-1} = z+y)P(Y_{t-1} = y)}{P(X_{t-1} - Y_{t-1} = z)} = \frac{\mathrm{e}^{-\lambda_1-\lambda_2-z\theta-2y\theta}(\lambda_2,\theta)_y(\lambda_1,\theta)_{z+y}}{\sum_{y=0}^{\infty} \mathrm{e}^{-\lambda_1-\lambda_2-z\theta-2y\theta}(\lambda_2,\theta)_y(\lambda_1,\theta)_{z+y}},$$

where $(\lambda,\theta)_x = \frac{\lambda(\lambda+x\theta)^{x-1}}{x!}$.

Thus,

$$
\begin{aligned}
&E(Z_t^2|Z_{t-1}=z) \\
&= \sum_{y=0}^{\infty} \frac{P(X_{t-1}=z+y)P(Y_{t-1}=y)}{P(X_{t-1}-Y_{t-1}=z)} \{p[(z+y)(z+y(1-p))+y^2]-pq[W_{(z+y)}-W_{(y)}]\} \\
&= \frac{\sum_{y=0}^{\infty} \mathrm{e}^{-2y\theta}(\lambda_2,\theta)_y(\lambda_1,\theta)_{y+z}\{p[(z+y)(z+y(1-p))+y^2]-pq[W_{(z+y)}-W_{(y)}]\}}{\sum_{y=0}^{\infty} \mathrm{e}^{-2y\theta}(\lambda_2,\theta)_y(\lambda_1,\theta)_{z+y}}.
\end{aligned}
$$

Then, the conditional variance given $Z_{t-1}=z, z\geq 0$ is

$$
\begin{aligned}
&\quad \mathrm{Var}(Z_t|Z_{t-1}=z) = E(Z_t^2|Z_{t-1}=z) - [E(Z_t|Z_{t-1}=z)]^2 \\
&= \frac{\sum_{y=0}^{\infty} e^{-2y\theta}(\lambda_2,\theta)_y(\lambda_1,\theta)_{y+z}\{p[(z+y)(z+y(1-p))+y^2-q[W_{(z+y)}+W_{(y)}]\}}{\sum_{y=0}^{\infty} e^{-2y\theta}(\lambda_2,\theta)_y(\lambda_1,\theta)_{z+y}} \\
&- \quad (pZ_{t-1}+\mu_\nu)^2.
\end{aligned}
$$

For $z<0$ we have

$$
\begin{aligned}
&E(Z_t^2|Z_{t-1}=z) = E[(S_t(X_{t-1})-S_t(Y_{t-1}))^2|X_{t-1}-Y_{t-1}=z] \\
&= \sum_{x=0}^{\infty} \frac{P(X_{t-1}=x)P(Y_{t-1}=x-z)}{P(X_{t-1}-Y_{t-1}=z)} E[((S_t(X_{t-1})-S_t(Y_{t-1}))^2|X_{t-1}=x, Y_{t-1}=x-z)].
\end{aligned}
$$

But

$$
\begin{aligned}
&E[(S_t(X_{t-1})-S_t(Y_{t-1})^2|X_{t-1}=x, Y_{t-1}=x-z)] \\
&= E(S_t^2(X_{t-1})|X_{t-1}=x, Y_{t-1}=x-z)) + E(S_t^2(Y_{t-1})|X_{t-1}=x, Y_{t-1}=x-z)) \\
&-E[S_t(X_{t-1})|X_{t-1}=x, Y_{t-1}=x-z)]E[S_t(Y_{t-1})|X_{t-1}=x, Y_{t-1}=x-z)]. \qquad (2.16)
\end{aligned}
$$

We know that if the random variable U has quasi-binomial distribution then Var(U) is given by (2.2), so

$$
\begin{aligned}
E(S_t^2(X_{t-1})|X_{t-1}=x) &= \mathrm{Var}(S_t(X_{t-1})|X_{t-1}=x) + E^2(S_t(X_{t-1})|X_{t-1}=x) \\
&= p[x^2 - qW_{(x)}] \qquad (2.17)
\end{aligned}
$$

and

$$
E(S_t^2(Y_{t-1})|Y_{t-1}=x-z) = p[(x-z)^2 - qW_{(x-z)}], \qquad (2.18)
$$

where $W_{(x)} = \sum_{j=0}^{x-2} \frac{x!\theta^j}{(x-j-2)!(1+x\theta)^{j+1}}$. X_{t-1} and Y_{t-1} are independent random variable, therefore

$$
\begin{aligned}
&E[S_t(X_{t-1})|X_{t-1}=x, Y_{t-1}=x-z)]E[S_t(Y_{t-1})|X_{t-1}=x, Y_{t-1}=x-z)] \\
&= x(x-z)p^2. \qquad (2.19)
\end{aligned}
$$

Replacing (2.17), (2.18) and (2.19) in (2.16), we obtain

$$E(Z_t^2|Z_{t-1}=z)$$
$$=\sum_{x=0}^{\infty}\frac{P(X_{t-1}=x)P(Y_{t-1}=x-z)}{P(X_{t-1}-Y_{t-1}=z)}\{p[x^2-qW_{(x)}]+p[(x-z)^2-qW_{(x-z)}]-x(x-z)p^2\}$$
$$=\sum_{x=0}^{\infty}\frac{P(X_{t-1}=x)P(Y_{t-1}=x-z)}{P(X_{t-1}-Y_{t-1}=z)}\{p[(x-z)(x(1-p)-z)+x^2-q(W_{(x)}+W_{(x-z)})]\}.$$

We have that

$$\frac{P(X_{t-1}=x)P(Y_{t-1}=x-z)}{P(X_{t-1}-Y_{t-1}=z)}=\frac{\mathrm{e}^{-\lambda_1-\lambda_2+z\theta-2x\theta}(\lambda_2,\theta)_{(x-z)}(\lambda_1,\theta)_x}{\sum_{y=0}^{\infty}\mathrm{e}^{-\lambda_1-\lambda_2-z\theta-2y\theta}(\lambda_2,\theta)_y(\lambda_1,\theta)_{y+z}}$$

where $(\lambda,\theta)_x=\frac{\lambda(\lambda+x\theta)^{x-1}}{x!}$. Then,

$$E(Z_t^2|Z_{t-1}=z)$$
$$=\sum_{x=0}^{\infty}\frac{P(X_{t-1}=z+y)P(Y_{t-1}=y)}{P(X_{t-1}-Y_{t-1}=z)}\{p[(x)^2-qW_{(x)}]+p((x-z)^2-qW_{(x-z)})-x(x-z)p^2\}$$
$$=\frac{\sum_{x=0}^{\infty}\mathrm{e}^{-2x\theta}(\lambda_2,\theta)_{(x-z)}(\lambda_1,\theta)_x\{p[(x-z)(x(1-p)-z)+x^2-q(W_{(x)}+W_{(x-z)})]\}}{\sum_{y=0}^{\infty}\mathrm{e}^{-2(y+z)\theta}(\lambda_2,\theta)_y(\lambda_1,\theta)_{y+z}}$$

Thus, the conditional variance given $Z_{t-1}=z, z<0$ is

$$\mathrm{Var}(Z_t|Z_{t-1}=z)=E(Z_t^2|Z_{t-1}=z)-[E(Z_t|Z_{t-1}=z)]^2$$
$$=\frac{\sum_{x=0}^{\infty}(\lambda_2,\theta)_{(x-z)}(\lambda_1,\theta)_x\{p[(x-z)(x(1-p)-z)+x^2-q(W_{(x)}+W_{(x-z)})]\}\mathrm{e}^{-2x\theta}}{\sum_{y=0}^{\infty}(\lambda_2,\theta)_y(\lambda_1,\theta)_{y+z}\mathrm{e}^{-2(y+z)\theta}}$$
$$-(pZ_{t-1}+\mu_\nu)^2.$$

3 Improved estimation for the NGINAR process

Resumo

Neste capítulo, desenvolvemos estimadores aproximadamente não-viesados para um dos parâmetros do processo autoregressivo de valores inteiros de primeira ordem com distribuição marginal geométrica baseado no operador thinning binomial negativo, processo NGINAR(1). Consideramos os estimadores de Yule-Walker do processo, derivamos o termo de primeira ordem do viés para um dos parâmetros do modelo e propomos um novo estimador com viés ajustado. Um estudo de simulação de Monte Carlo é considerado para analisar o comportamento do novo estimador.

Palavras-chave: Correção de viés. Processo NGINAR(1). Estimadores de mínimos quadrados condicionais. Estimadores de Yule-Walker.

Abstract

In this paper, we developed nearly unbiased estimators for the parameters of the first-order integer-valued autoregressive process with geometric marginal distributions based on the negative binomial thinning, NGINAR(1) process. We consider the Yule-Walker estimators, derive the first order term of bias for one of the parameters and propose a new bias-adjusted estimator. Monte Carlo simulation study is considered to analyse the behaviour of the new estimator.

Keywords: Bias correction. Conditional least square estimator. NGINAR(1) process. Yule-Walker estimator.

3.1 Introduction

Integer-valued time series have received a lot of attention in recent decades. The motivation for this was the need for new models that apply to count series whose observations are correlated. McKenzie (1985) and Alzaid and Al-Osh(1987), independently, were the pioneered to develop a model to time series of integer values based on binomial thinning operator (Steutel and Van Harn, 1979). They introduced the first-order non-negative integer-valued autoregressive (INAR(1)) process with Poisson innovations and Poisson marginal distribution to modelling discrete-valued time series.

McKenzie (1985) and Al-Osh and Aly (1992) proposed INAR(1) models with negative binomial and geometric marginals based on the binomial thinning operator. Ristić et al. (2009) defined the negative binomial thinning operator using geometric counting series and introduced the new stationary first-order integer-valued autoregressive process with geometric marginals (NGINAR(1)). The authors obtained properties of the negative binomial thinning operator and several properties of the process. Estimators of the unknown parameters of the model were obtained by conditional least squares (CLS), as well as a Yule-Walker (YW) and the maximum likelihood (ML) methods, asymptotic properties of the estimators are presented, simulation study and an application to real data set to show the applicability of the NGINAR(1) model. Bakouch (2010) obtained expressions for the factorial and joint higher-order moments and cumulants, the spectral and bispectral densities for NGINAR(1) process. Awale et al. (2017) compared coherent forecasts of two INAR models with geometric marginal, GINAR(1) model based on the binomial thinning operator and NGINAR(1) model based on the negative binomial thinning operator and discuss two applications of geometric INAR models to epidemiological data.

The NGINAR(1) model is interesting in sense that the Poisson INAR(1) (Alzaid and Al-Osh, 1987) is not always adequate on real situations since its mean and variance are equal, when the mean is smaller than the variance (overdispersed data set) the NGINAR(1) is more adequate than the Poisson INAR(1).

Although researchers have found parameters estimators of the time series models that have good asymptotic properties, sometimes there is a problem when they work with a small number of observations. The traditional approaches to parameter estimation in integer-valued time series, YW, CLS and ML, are biased in finite samples. Some authors in order to try to minimize this problem have taken as an alternative the bias correction of estimators. For instance, Jung et. al (2005) considered estimation in AR(1) models with discrete support and the use of bias correction techniques is applied to simulated Poisson INAR(1) processes. Weiß and Kim (2013) considered the binomial AR(1) process, presented four approaches for estimating its model parameters, investigated the finite-sample performance of the estimators in a simulation study and analyze the effect of applying a 2-block jackknife. Bourguignon and Vasconcellos (2015a) derived the second-order bias of the SD estimator of the Poisson INAR(1) process for one of the parameters and defined a bias-reduced estimator. Weiss and Schweer (2016) considered Poisson INAR(1) and INARCH(1) count data time series, explicit asymptotic approximations for bias and standard deviation of common moment estimators are derived and their finite-sample performance is shown by simulations.

The purpose of this chapter is, considering the NGINAR(1) process, to derive a first-order term of bias of the YW estimator for one of the parameters based on Weiß and Schweer (2016), so to obtain a bias reduced estimator. To that end, we obtain some asymptotic results using the NGINAR(1) process already demonstrated in the paper by Weiß and Kim (2013) for the Poisson INAR(1) and Poisson INARCH(1) (Heinen, 2003; Weiß 2010) processes.

The chapter unfolds as follows. In Section 2, we introduced the NGINAR(1) process and some basic properties are outlined and the moment estimators are presented. In Section 3, we derive the expression of the first-order term of bias of the YW estimator. Numerical results from Monte Carlo simulations are presented and discussed in Section 4. Finally, Section 5 concludes the chapter.

3.2 NGINAR(1) process

Ristić et al. (2009) introduced the NGINAR(1) process based on the negative binomial thinning operator. Let $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. The NGINAR(1) process is given by

$$X_t = \alpha * X_{t-1} + \epsilon_t, \ t \in \mathbb{N}, \tag{3.1}$$

where the operator $*$ is defined as $\alpha * X = \sum_{i=1}^{X} Y_i$, $\alpha \in [0, 1)$, $\{X_t\}_{\mathbb{N}_0}$ is a stationary process with geometric $(\mu/(1+\mu))$ marginals, with probability mass function given by $\Pr(X_t = x) = \mu^x/(1+\mu)^{x+1}$, $x \in \mathbb{N}_0$, $\{Y_i\}$ is a sequence of independent identically distributed (i.i.d.) random variables with geometric$(\alpha/(1+\alpha))$ distribution, $\{\epsilon_t\}$ is a sequence of i.i.d. random variables independent of $\{Y_i\}$ and X_{t-l} and ϵ_t are independent for all $l \geq 1$. The unconditional mean and variance of X_t following the NGINAR(1) process are given by $\mu_X = \mathrm{E}(X_t) = \mu$ and $\sigma_X^2 = \mathrm{Var}(X_t) = \mu(1+\mu)$, respectively.

Ristić et al. (2009) proved that the random variable ϵ_t is a mixture of two random variables with geometric$(\mu/(1+\mu))$ and geometric$(\alpha/(1+\alpha))$ and the probability mass function of ϵ_t is given by

$$\Pr(\epsilon_t = l) = \left(1 - \frac{\alpha\mu}{\mu - \alpha}\right) \frac{\mu^l}{(1+\mu)^{l+1}} + \frac{\alpha\mu}{\mu - \alpha} \frac{\alpha^l}{(1+\alpha)^{l+1}}, \quad l \in \mathbb{N},$$

where $\alpha \in [0, \mu/(1+\mu)]$, so that the function is well defined. The unconditional mean and variance of ϵ_t are given by $\mathrm{E}(\epsilon_t) = (1-\alpha)\mu$ and $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_t) = \mu(1+\alpha)((1+\mu)(1-\alpha) - \alpha)$.

Let $\{X_t\}_{\mathbb{N}_0}$ be a process following the NGINAR(1) model. Then, X_t follows a strictly stationary ergodic Markov chain with transition probabilities

$$\Pr(X_t = k|X_{t-1} = 0) = \left(1 - \frac{\alpha\mu}{\mu - \alpha}\right) \frac{\mu^k}{(1+\mu)^{k+1}} + \frac{\alpha\mu}{\mu - \alpha} \frac{\alpha^k}{(1+\alpha)^{k+1}}, \quad k \in \mathbb{N},$$

$$p_{k|l} = \frac{\mu\alpha^{k+1}}{(\mu-\alpha)(1+\alpha)^{k+l+1}} \binom{l+k}{k} + \left(1 - \frac{\alpha\mu}{\mu-\alpha}\right) \frac{\mu^k}{(1+\alpha)^l (1+\mu)^{k+1}} \sum_{m=0}^{k} \binom{l+m-1}{l-1} \left[\frac{\alpha(1+\mu)}{\mu(1+\alpha)}\right]^m, l \geq 1,$$

where $\binom{\cdot}{\cdot}$ is the standard combinatorial symbol and $p_{k|l} := \Pr(X_t = k|X_{t-1} = l)$.

The autocorrelation function of the NGINAR(1) process is given by $\rho(k) = \alpha^k, \alpha \in [0, \mu/(1+\mu)], k \geq 0$, which has exponential decay, as in the classical Gaussian AR(1) model.

The expressions of k-step ahead conditional mean and variance are given, respectively, by

$$E(X_{t+k}|X_t) = \alpha^k X_t + \frac{1-\alpha^k}{1-\alpha}\mu_\epsilon,$$

and

$$\mathrm{Var}(X_{t+k}|X_t) =$$

$$\frac{\alpha^k(1+\alpha)(1-\alpha^k)}{1-\alpha} X_t + \frac{1-\alpha^{2k}}{1-\alpha^2}\sigma_\epsilon^2 + \frac{\alpha(1+\alpha+2\mu_\epsilon)}{1-\alpha}\left(\frac{1-\alpha^{2k}}{1-\alpha^2} - \alpha^{k-1}\frac{1-\alpha^k}{1-\alpha}\right)\mu_\epsilon$$

$$+\left[\frac{1-\alpha^{2k}}{1-\alpha^2} - \frac{1-2\alpha^k+\alpha^{2k}}{(1-\alpha)^2}\right]\mu_\epsilon^2.$$

Bakouch (2010) considered the NGINAR(1) process and derived expressions for the joint higher order moments and cumulants of the process. Such expressions are essential for the development of the properties of the parameter estimators of the NGINAR(1) model. Denote the kth-order joint moment of the random variables $X_t, X_{t+s_1}, \ldots, X_{t+s_{k-1}}$ by

$$E(X_t \cdot X_{t+s_1} \cdot \ldots \cdot X_{t+s_{k-1}}) = \mu(s_1, \ldots, s_{k-1}).$$

Formulae for second-order, third-order and fourth-order joint moments of the NGINAR(1) process by Bakouch (2010) are given below.

Theorem 3.2.1. *Let $(X_t)_{\mathbb{N}_0}$ be a stationary NGINAR(1) process as (3.1).*

(i) The second-order joint moment of the NGINAR(1) process is given by

$$\mu(s) = \mu(1+\mu)\alpha^s + \mu^2, \; s \geq 0.$$

(ii) For the third-order joint moment of the NGINAR(1), we have that

$$\begin{aligned}
\mu(0,s) &= \mu(1+5\mu+4\mu^2)\alpha^s + \mu^2(1+2\mu),\ s \geq 0,\\
\mu(s,s) &= \tfrac{2\mu(1+\mu)^2}{1-\alpha}\alpha^{2s+1} + \tfrac{2\mu^2(1+\mu)}{1-\alpha}\alpha^{2s} + \tfrac{\mu(1+\mu)(1-2\mu)}{1-\alpha}\alpha^{s+1} + \tfrac{\mu(1+\mu)(1+2\mu)}{1-\alpha}\alpha^s\\
&+ \mu^2(1+2\mu),\ s \geq 0,\\
\mu(s,u) &= \alpha^{u-s}\mu(s,s) + (1-\alpha^{u-s})\mu\mu(s), u > s.
\end{aligned}$$

(iii) For the fourth-order joint moment of the NGINAR(1) process, we have that

$$\begin{aligned}
\mu(0,0,s) &= \mu(1+13\mu+30\mu^2+18\mu^3)\alpha^s + \mu^2(1+6\mu+6\mu^2), s \geq 0\\
\mu(0,s,s) &= \alpha^2\mu(0,s-1,s-1) + \alpha(1+\alpha+2(1-\alpha)\mu)\mu(0,s-1)\\
&\quad + (\sigma_\epsilon^2+\mu_\epsilon^2)\mu(0),\ s > 0,
\end{aligned}$$

$$\begin{aligned}
\mu(0,s,u) &= \alpha\mu(0,s,u-1) + (1-\alpha)\mu\mu(0,s),\ u > s > 0,\\
\mu(s,u,v) &= \alpha\mu(s,u,v-1) + (1-\alpha)\mu\mu(s,u),\ v > u > s > 0,\\
\mu(s,u,u) &= \alpha^2\mu(s,u-1,u-1) + \alpha(1+\alpha+2(1-\alpha)\mu)\mu(s,u-1) + (\sigma_\epsilon^2+\mu_\epsilon^2),\\
&\quad u > s > 0,\\
\mu(s,s,u) &= \alpha\mu(s,s,u-1) + (1-\alpha)\mu\mu(s,s),\ u > s > 0,\\
\mu(s,s,s) &= \alpha^3\mu(s-1,s-1,s-1) + 3\alpha^2(1+\alpha+(1-\alpha)\mu)\mu(s-1,s-1)\\
&\quad +\alpha(-6\alpha\mu^2+6\mu^2-9\alpha^2\mu-3\alpha\mu+6\mu+2\alpha^2+3\alpha+1)\mu(s-1)\\
&\quad +\mu^2(-6\alpha^3-6\mu\alpha^2-6\alpha^2-6\mu^2\alpha-6\mu\alpha-\alpha+6\mu^2+6\mu+1),\ s > 0.
\end{aligned}$$

In particular,

$$\begin{aligned}
E(X_t^2) &= \mu(0) = \mu(1+2\mu),\ E(X_t^3) = \mu(0,0) = \mu(1+6\mu+6\mu^2),\\
E(X_t^4) &= \mu(0,0,0) = \mu(1+14\mu+36\mu^2+24\mu^3).
\end{aligned}$$

From Theorem 3.2.1, we obtain that the closed form for $\mu(0,k,k)$ and $\mu(1,k,k)$ are given by

$$\begin{aligned}
\mu(0,k,k) &= \alpha^{2k}\mu(0,0,0) + \alpha(1+\alpha+2(1-\alpha)\mu)\left[\frac{\mu(1+\mu)(1+4\mu)}{1-\alpha}(\alpha^{k-1}-\alpha^{2k-1})\right.\\
&\quad \left. -\mu\mu(0)\frac{\alpha^{2k}}{1-\alpha^2}\right] - (\sigma_\epsilon^2+\mu_\epsilon^2)\mu(0)\frac{\alpha^{2k}}{1-\alpha^2} + \mu(0)^2, \qquad (3.2)
\end{aligned}$$

and

$$\begin{aligned}\mu(1,k,k) &= \alpha^{2(k-1)}\mu(1,1,1) + \alpha(1+\alpha+2(1-\alpha)\mu)\,[\frac{\mu(1,1)}{1-\alpha}(\alpha^{k-2}-\alpha^{2k-3}) \\ &\quad -\mu\mu(1)(\frac{\alpha^{2k-2}}{1-\alpha^2}+\frac{\alpha^{k-2}-\alpha^{2k-3}}{1-\alpha})] - (\sigma_\epsilon^2+\mu_\epsilon^2)\mu(1)\frac{\alpha^{2k-2}}{1-\alpha^2} \\ &\quad +\mu(0)\mu(1). \end{aligned} \tag{3.3}$$

3.2.1 *Estimation of the unknown parameters*

In real situations, it is not possible to know the true values of the model parameters, so it is needed to estimate them from a given time series data. In the NGINAR(1) model we are interested in estimating the unknown parameter vector $\eta = (\alpha, \mu)^\top$. Estimation of the NGINAR(1) parameters was considered by Ristić et al. (2009), who present three approaches for estimating its model parameters based on a given time series data and derived expressions for the asymptotic distributions of these estimators. The approaches were CLS estimation, YW estimation or moments estimation and, finally, maximum likelihood (ML) estimation. Here, we present only the ML and Yule-Walker approaches.

Among the most used estimators in the literature is ML. The ML estimator is obtained by maximizing the log-likelihood function $\ell(\alpha, \mu)$. The log-likelihood for $X_1, X_2, \ldots, X_T$ stem from a stationary NGINAR(1) process is given by

$$\ell(\alpha,\mu) = x_1\log(\mu) - (x_1+1)\log(1+\mu) + \sum_{t=2}^{T}\log[P(X_t=k|X_{t-1}=l)].$$

where the transition probabilities $P(X_t = k|X_{t-1} = l)$ are given by (3.2).

In general, for time series is naturally consider the CML instead the ML estimator when the marginal distribution is unknown. The CML estimators are obtained by maximizing the conditional log-likelihood function, for $X_1, X_2, \ldots, X_T$ stem from a stationary NGINAR(1) process, we have the conditional log-likelihood is given by

$$\ell_c(\alpha,\mu) = \log\left[\prod_{t=2}^{T}P(X_t=k|X_{t-1}=l)\right] = \sum_{t=2}^{T}\log[P(X_t=k|X_{t-1}=l)],$$

In general, to obtain the maximum value of the log-likelihood and conditional log-likelihood function is necessary a numerical method. The YW estimator is a reasonable alternative to ML and CML, because has a simple expression.

Let $X_1, \ldots, X_T$ be a realization from a NGINAR(1) process with parameters α and μ. Let $\bar{X}_T = \frac{1}{T}\sum_{i=1}^{T} X_i$ be the sample mean. We have that $\widehat{\mu} = \bar{X}_T$ is an unbiased estimator of μ and knowing that the autocorrelation function is given by $\rho(k) = \alpha^k$, YW estimator of α based on the empirical first-order autocorrelation is given by the expression

$$\widehat{\alpha}_{YW} = \widehat{\rho}(1) = \frac{\sum_{t=2}^{T}(X_t - \overline{X}_T)(X_{t-1} - \overline{X}_T)}{\sum_{t=1}^{T}(X_t - \overline{X}_T)^2}.$$

The asymptotic distribution of YW estimators is given in next proposition.

Proposition 3.2.2. *(Ristić et al., 2009) The joint distribution of the YW estimators satisfies*

$$\sqrt{T}(\widehat{\alpha}_{YW} - \alpha, \widehat{\mu} - \mu)^\top \xrightarrow{d} N(0, \mathbf{\Sigma}_{\alpha,\mu}).$$

where

$$\mathbf{\Sigma}_{\alpha,\mu} = \begin{pmatrix} \dfrac{(1+\alpha)(\mu + \mu^2 + \alpha + \alpha\mu - \alpha\mu^2)}{\mu(1+\mu)} & \dfrac{\alpha(1+\alpha)}{1-\alpha} \\ \\ \dfrac{\alpha(1+\alpha)}{1-\alpha} & \dfrac{\mu(1+\mu)(1+\alpha)}{1-\alpha} \end{pmatrix}$$

where $\xrightarrow{d}$ *meaning convergence in distribution.*

Ristić et al. (2009) showed that the YW estimators have same asymptotic distribution of CLS estimators.

3.3 Bias correction Yule-Walker estimator

In general the classical approaches of estimation produce biased estimators, but this is not a problem when the sample size is large. However, for moderate sample sizes, bias can become a problem. When the sample size is small it becomes necessary to use a bias correction to obtain a reduced-bias estimator. In this section, we will derive the first-order bias of the YW estimator of α and, based on Weiß and Schweer (2016), we propose a bias-corrected YW estimator.

We have that the NGINAR(1) process is a strictly stationary and ergodic process (Ristić et al. 2009). In the same way of Weiß and Schweer (2016), to obtain the bias correction to $\widehat{\alpha}$ estimator we need the following result whose proof is in the Appendix.

Proposition 3.3.1. *Let* $(X_t)_{\mathbb{N}_0}$ *be a stationary NGINAR(1) process as (3.1). Define the trivariate process* ξ_t *by*

$$\xi_t = (X_t - \mu, X_t^2 - \mu(0), X_t X_{t-1} - \mu(1))^\top, \quad \textit{with} \quad \mu(k) := E(X_t X_{t-k}). \tag{3.4}$$

Then,

$$T^{-1/2}\sum_{t=1}^{T}\xi_t \xrightarrow{d} N(0, \Sigma),$$

with $\Sigma = (\sigma_{i,j})$ *being*

$$\sigma_{i,j} = E(\xi_0^i \xi_0^j) + \sum_{k=1}^{\infty} E(\xi_0^i \xi_k^j) + \sum_{k=1}^{\infty} E(\xi_k^i \xi_0^j),$$

ξ_t^h *representing the* h*th component of* ξ_t *and the* $\sigma_{i,j}$ *are given by*

$$\begin{aligned}
\sigma_{1,1} &= \frac{1+\alpha}{1-\alpha}\,\sigma_X^2, \;\; \sigma_{1,2} = \frac{1+\alpha+\alpha^2-\alpha^3}{(1-\alpha)(1-\alpha^2)}\,\sigma_X^2 + \frac{2+3\alpha+2\alpha^2}{1-\alpha^2}\mu\sigma_X^2,\\
\sigma_{1,3} &= \frac{\alpha(2+2\alpha-2\alpha^3)}{(1-\alpha)(1-\alpha^2)}\,\sigma_X^2 + \frac{2(1+3\alpha+2\alpha^2+\alpha^3)}{1-\alpha^2}\,\mu\sigma_X^2, \;\; \sigma_{2,1} = \sigma_{1,2},\\
\sigma_{2,2} &= \frac{1+\alpha+3\alpha^2-\alpha^3}{(1-\alpha)(1-\alpha^2)}\sigma_X^2 + \frac{1+11\alpha+5\alpha^2-\alpha^3}{(1-\alpha)(1-\alpha^2)}\mu\sigma_X^2 + \frac{9+16\alpha+9\alpha^2}{1-\alpha^2}\mu^2\sigma_X^2\\
&\quad + \frac{11(1+\alpha^2)}{1-\alpha^2}\sigma_X^4,\\
\sigma_{2,3} &= \frac{(2+3\alpha+4\alpha^2-\alpha^3)}{(1-\alpha)(1-\alpha^2)}\alpha\sigma_X^2 + \frac{(2+21\alpha+4\alpha^2-11\alpha^3)}{(1-\alpha)(1-\alpha^2)}\alpha\mu\sigma_X^2\\
&\quad + \frac{(8+10\alpha+19\alpha^2-2\alpha^3)}{1-\alpha^2}\mu^2\sigma_X^2 + \frac{(6\alpha-3\alpha^2+12\alpha^3)}{1-\alpha^2}\sigma_X^4 + \frac{3\alpha^2(1+\alpha)\mu}{1-\alpha^2}\\
&\quad + \frac{(1+6\alpha^2-6\alpha^3)}{1-\alpha^2}\mu^2 - \frac{9\alpha^3}{1-\alpha^2}\mu^3 - \frac{1+3\alpha^2}{1-\alpha^2}\mu^4,\\
\sigma_{3,1} &= \sigma_{1,3}, \quad \sigma_{3,2} = \sigma_{2,3},\\
\sigma_{3,3} &= \frac{\alpha+2\alpha^2+10\alpha^3+5\alpha^4-6\alpha^5}{(1-\alpha)(1-\alpha^2)}\sigma_X^2 + \frac{1+19\alpha^2-6\alpha^3+14\alpha^4}{1-\alpha^2}\sigma_X^4\\
&+ \frac{6\alpha+7\alpha^2+25\alpha^3-21\alpha^4-\alpha^5}{(1-\alpha)(1-\alpha^2)}\mu\sigma_X^2 + \frac{4+20\alpha+11\alpha^2-15\alpha^3+25\alpha^4-43\alpha^5}{(1-\alpha)(1-\alpha^2)}\mu^2\sigma_X^2\\
&+ \frac{\alpha^2}{1-\alpha^2}\mu + \frac{4\alpha-\alpha^2}{1-\alpha^2}\mu^2 + \frac{2}{1-\alpha^2}\mu^4.
\end{aligned}$$

To obtain the bias correction we will proceed as Weiß and Schweer (2016). Define the function $f : \mathbb{R}^3 \to \mathbb{R}$:

$$f(x_1, x_2, x_3) := \frac{x_3 - x_1^2}{x_2 - x_1^2}.$$

Considering $\eta = (\mu_X, \mu(0), \mu(1))$, where μ_X is the mean of the process and $\mu(k)$ as defined in (3.4), and the vector

$$\mathbf{X_T} = \left(\frac{1}{T}\sum_{t=1}^{T} X_t, \frac{1}{T}\sum_{t=1}^{T} {X_t}^2, \frac{1}{T}\sum_{t=1}^{T} X_t X_{t-1}\right),$$

we have that

$$f(\eta) = \rho(1) \quad \text{and} \quad f(\mathbf{X_T}) = \frac{\widehat{\gamma}(1)^*}{\widehat{\gamma}(0)^*},$$

where $\widehat{\gamma}(0)^* = \widehat{\gamma}(0) = \frac{1}{T}\sum_{t=1}^{T} X_t^2 - \bar{X}_T^2$, $\widehat{\gamma}(1)^* = \widehat{\gamma}(1) + \frac{1}{T}[\bar{X}_T(X_T - \bar{X}_T) + X_1(\bar{X}_T - X_0)]$ and $\widehat{\gamma}(k) = \frac{1}{T}\sum_{t=1}^{T-k}(X_t - \bar{X}_T)(X_{t+k} - \bar{X}_T)$ is the sample autocovariance function.

To compute the bias correction we need the second-order partial derivatives of the f function to find the Hessian matrix, so we use a second-order Taylor expansion for f function. The first-order partial derivatives are

$$\frac{\partial f}{\partial x_1} = \frac{2x_1(x_3 - x_2)}{(x_2 - x_1^2)^2}, \quad \frac{\partial f}{\partial x_2} = \frac{x_1^2 - x_3}{(x_2 - x_1^2)^2}, \quad \frac{\partial f}{\partial x_3} = \frac{1}{x_2 - x_1^2},$$

and the Hessian matrix is given by

$$\mathbf{H}_f(x_1, x_2, x_3) = \begin{pmatrix} \dfrac{2(x_3 - x_2)(x_2 + 3x_1^2)}{(x_2 - x_1^2)^3} & \dfrac{-2x_1(2x_3 - x_2 - x_1^2)}{(x_2 - x_1^2)^3} & \dfrac{2x_1}{(x_2 - x_1^2)^2} \\ * & \dfrac{2(x_3 - x_1^2)}{(x_2 - x_1^2)^3} & \dfrac{-1}{(x_2 - x_1^2)^2} \\ * & * & 0 \end{pmatrix}.$$

We have that $(\mu_X, \mu(0), \mu(1)) = (\mu, \mu(1 + 2\mu), \mu(\mu + \alpha(1 + \mu)))$, then

$$\mathbf{H}_f(\mu_X, \mu(0), \mu(1)) = \begin{pmatrix} \dfrac{-2(1-\alpha)(\sigma_X^2 + 4\mu^2)}{\sigma_X^4} & \dfrac{2(1-2\alpha)\mu}{\sigma_X^4} & \dfrac{2\mu}{\sigma_X^4} \\ * & \dfrac{2\alpha}{\sigma_X^4} & \dfrac{-1}{\sigma_X^4} \\ * & * & 0 \end{pmatrix}.$$

Now, denote the vector $\mathbf{Y}_T := \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_t$. By a second order Taylor expansion for f function, we obtain the bias correction from

$$\begin{aligned} B(\widehat{\alpha}_{YW}) \;&:=\; T\,E[\widehat{\rho}(1) - \rho(1)] \approx E\left[\frac{1}{2}\mathbf{Y}_T^\top \mathbf{H}_f(\mu, \mu(0), \mu(1))\mathbf{Y}_T\right] \\ &\approx\; \frac{1}{2}(h_{1,1}\sigma_{1,1} + h_{2,2}\sigma_{2,2} + 2h_{1,2}\sigma_{1,2} + 2h_{1,3}\sigma_{1,3} + 2h_{2,3}\sigma_{2,3}), \end{aligned}$$

where $h_{i,j}$ representing the (i,j) element from $\mathbf{H}_f$, the Hessian matrix. Thus,

$$B(\widehat{\alpha}_{YW}) = \frac{1}{2}\left[\frac{-2(1-\alpha)(\sigma_X^2+4\mu^2)}{\sigma_X^4}\sigma_{1,1} + \frac{4\alpha}{\sigma_X^4}\sigma_{2,2} + \frac{4(1-2\alpha)\mu}{\sigma_X^4}\sigma_{1,2} + \frac{4\mu}{\sigma_X^4}\sigma_{1,3} - \frac{2}{\sigma_X^4}\sigma_{2,3}\right].$$

Replacing the terms of covariance matrix given in Proposition 3.3.1, we have

$$\begin{aligned} B(\widehat{\alpha}_{YW}) &= \frac{1}{\sigma_X^4}\Bigg\{-(1-\alpha)(\sigma_X^2+4\mu^2)\frac{1+\alpha}{1-\alpha}\,\sigma_X^2 + 2\alpha\Bigg[\frac{1+\alpha+3\alpha^2-\alpha^3}{(1-\alpha)(1-\alpha^2)}\sigma_X^2 \\ &+ \frac{11(1+\alpha^2)}{1-\alpha^2}\sigma_X^4 + \frac{1+11\alpha+5\alpha^2-\alpha^3}{(1-\alpha)(1-\alpha^2)}\mu\sigma_X^2 + \frac{9+16\alpha+9\alpha^2}{1-\alpha^2}\mu^2\sigma_X^2\Bigg] \\ &+ 2(1-2\alpha)\mu\Bigg[\frac{2+3\alpha+2\alpha^2}{1-\alpha^2}\mu\sigma_X^2 + \frac{1+\alpha+\alpha^2-\alpha^3}{(1-\alpha)(1-\alpha^2)}\,\sigma_X^2\Bigg] \\ &+ 2\mu\Bigg[\frac{\alpha(2+2\alpha-2\alpha^3)}{(1-\alpha)(1-\alpha^2)}\,\sigma_X^2 + \frac{2(1+3\alpha+2\alpha^2+\alpha^3)}{1-\alpha^2}\,\mu\sigma_X^2\Bigg] \\ &- 2\Bigg[\frac{(2+3\alpha+4\alpha^2-\alpha^3)}{(1-\alpha)(1-\alpha^2)}\alpha\sigma_X^2 + \frac{(2+21\alpha+4\alpha^2-11\alpha^3)}{(1-\alpha)(1-\alpha^2)}\alpha\mu\sigma_X^2 \\ &+ \frac{(8+10\alpha+19\alpha^2-2\alpha^3)}{1-\alpha^2}\mu^2\sigma_X^2 + \frac{(6\alpha-3\alpha^2+12\alpha^3)}{1-\alpha^2}\sigma_X^4 \\ &+ \frac{3\alpha^2(1+\alpha)\mu}{1-\alpha^2} + \frac{(1+6\alpha^2-6\alpha^3)}{1-\alpha^2}\mu^2 - \frac{9\alpha^3}{1-\alpha^2}\mu^3 - \frac{1+3\alpha^2}{1-\alpha^2}\mu^4\Bigg]\Bigg\}. \end{aligned}$$

After some algebra, we have

$$\begin{aligned} B(\widehat{\alpha}_{YW}) &= \frac{1}{\sigma_X^4}\Bigg[\frac{2(1-9\alpha^2-2\alpha^3+10\alpha^4)}{(1-\alpha)(1-\alpha^2)}\mu\sigma_X^2 + \frac{2(-8+\alpha^2+11\alpha^3+2\alpha^4)}{1-\alpha^2}\mu^2\sigma_X^2 \\ &+ \frac{(-2-4\alpha-2\alpha^2)\alpha}{(1-\alpha)(1-\alpha^2)}\sigma_X^2 + \frac{-1+9\alpha+7\alpha^2-\alpha^3}{1-\alpha^2}\sigma_X^4 - \frac{6\alpha^2(1+\alpha)}{1-\alpha^2}\mu \\ &- \frac{2(1+6\alpha^2-6\alpha^3)}{1-\alpha^2}\mu^2 + \frac{18\alpha^3}{1-\alpha^2}\mu^3 + \frac{2(1+3\alpha^2)}{1-\alpha^2}\mu^4\Bigg]. \end{aligned}$$

Finally, the bias is given by

$$\begin{aligned} B(\widehat{\alpha}_{YW}) &= \frac{-1+9\alpha+7\alpha^2-\alpha^3}{1-\alpha^2} + \frac{1}{\sigma_X^2}\Bigg\{\frac{(-2-10\alpha+4\alpha^2)\alpha}{(1-\alpha)(1-\alpha^2)} + \frac{2\mu}{1-\alpha^2}\big[-4\alpha^3-8\alpha^2-2\alpha \\ &+ (-7+4\alpha^2+1\alpha^3+2\alpha^4)\mu\big]\Bigg\} + \frac{6\alpha^3\mu^3}{(1-\alpha^2)\sigma_X^4}. \end{aligned} \tag{3.5}$$

Note that the bias depends on the α and μ values and it is important to remember that there is a dependence between α and μ ($\alpha \in [0, \mu/(1+\mu)]$), so the bias expression is extremely influenced by the μ value.

Using (3.5), we consider a bias-reduced YW estimator $\overline{\alpha}$ of α as

$$\overline{\alpha} = \widehat{\alpha}_{YW} - \frac{1}{T}B(\widehat{\alpha}_{YW}). \tag{3.6}$$

It is easy to see that $\overline{\alpha}$ has the same asymptotic normal distribution as the original estimator $\widehat{\alpha}_{YW}$ given in Proposition 3.2.2, since the first-order bias term is $O(T^{-1})$.

3.4 Numerical evaluation

In this section, we compare the finite-sample performances of the bias-reduced YW estimates, YW estimates and ML estimates. We performed a Monte Carlo simulation study with 5000 replications and sample sizes $T = 45, 60, 75$ and 100. The values of α in the simulation study were $\alpha = 0.1, 0.3 \text{ and } 0.4$ for a fixed $\mu = 1$ and we considered $\alpha = 0.2, 0.4 \text{ and } 0.5$ for a fixed $\mu = 4$. The choice of α values was made considering the restriction $\alpha \in [0, \mu/(1 + \mu)]$. The Monte Carlo simulation experiments were performed using the R programming language; see http://www.r-project.org.

Tables 3.1 and 3.2 present the simulated mean estimates and mean squared error (MSE) of the YW, bias-reduced YW (YW_c) and ML estimators. Based on both measures, we observe that, in general, the best estimator of α, in terms of bias, was our proposed estimator the bias-reduced YW estimator. Also, the YW estimator has the worst performance in terms of bias. In some cases, for instance, $T = 100$, the bias-reduced YW is not better than the ML estimator, but the estimates of both are very close. It is important remember that the bias-reduced estimator has a closed-form expression of easy implementation, where as the ML estimates have to be computed using a numerical optimization method. In terms of MSE we have that, in general, the ML estimators presents the smallest results, while the bias-reduced YW presents the largest results.

The approximations become better and the MSE become smaller as T increases, as expected. So, based on our numerical results we recommend the use of the reduced-bias YW as estimator of the α parameter of an NGINAR(1) process especially in small sample size case.

3.5 Concluding remarks

In this paper, we considered the NGINAR(1) process and developed nearly unbiased estimators for one of the unknown parameters. For this, we consider the Yule-Walker estimators, derive the second-order bias for the parameter and propose a new bias-adjusted estimator. A Monte Carlo simulation study is considered to analyse the behaviour of the new estimator. The

Table 3.1: Simulated mean for estimators and MSEs (in parentheses) of $\alpha = 0.1, 0.2$ and 0.3 fixed $\mu = 1$.

$\mu = 1$	$\alpha = 0.1$			$\alpha = 0.2$			$\alpha = 0.3$		
T	$\widehat{\alpha}_{\mathrm{YW}}$	$\widehat{\alpha}_{\mathrm{YWc}}$	$\widehat{\alpha}_{\mathrm{ML}}$	$\widehat{\alpha}_{\mathrm{YW}}$	$\widehat{\alpha}_{\mathrm{YWc}}$	$\widehat{\alpha}_{\mathrm{ML}}$	$\widehat{\alpha}_{\mathrm{YW}}$	$\widehat{\alpha}_{\mathrm{YWc}}$	$\widehat{\alpha}_{\mathrm{ML}}$
45	0.0656	0.1015	0.1321	0.1575	0.2122	0.2169	0.2380	0.3094	0.3171
	(0.0226)	(0.0328)	(0.0330)	(0.0247)	(0.0375)	(0.0432)	(0.0285)	(0.0426)	(0.0667)
60	0.0748	0.1013	0.1179	0.1650	0.2012	0.2042	0.2523	0.3005	0.3029
	(0.0172)	(0.0216)	(0.0171)	(0.0202)	(0.0247)	(0.0281)	(0.0210)	(0.0240)	(0.0428)
75	0.0785	0.0992	0.1129	0.1736	0.2029	0.2005	0.2616	0.2973	0.2962
	(0.0138)	(0.0161)	(0.0122)	(0.0153)	(0.0176)	(0.0180)	(0.0177)	(0.0182)	(0.0289)
100	0.0858	0.1010	0.1092	0.1803	0.2006	0.1988	0.2710	0.2960	0.2898
	(0.0105)	(0.0117)	(0.0081)	(0.0119)	(0.0127)	(0.0111)	(0.0135)	(0.0129)	(0.0160)

Table 3.2: Simulated mean for estimators and MSEs (in parentheses) of $\alpha = 0.2, 0.4$ and 0.5 fixed $\mu = 4$.

$\mu = 4$	$\alpha = 0.2$			$\alpha = 0.4$			$\alpha = 0.5$		
T	$\widehat{\alpha}_{\mathrm{YW}}$	$\widehat{\alpha}_{\mathrm{YWc}}$	$\widehat{\alpha}_{\mathrm{ML}}$	$\widehat{\alpha}_{\mathrm{YW}}$	$\widehat{\alpha}_{\mathrm{YWc}}$	$\widehat{\alpha}_{\mathrm{ML}}$	$\widehat{\alpha}_{\mathrm{YW}}$	$\widehat{\alpha}_{\mathrm{YWc}}$	$\widehat{\alpha}_{\mathrm{ML}}$
45	0.1633	0.2047	0.2238	0.3412	0.4092	0.3779	0.4250	0.5132	0.4496
	(0.0226)	(0.0287)	(0.0082)	(0.0237)	(0.0453)	(0.0102)	(0.0256)	(0.0766)	(0.0134)
60	0.1705	0.1998	0.2168	0.3538	0.3964	0.3829	0.4427	0.4941	0.4518
	(0.0162)	(0.0188)	(0.0063)	(0.0176)	(0.0196)	(0.0102)	(0.0183)	(0.0331)	(0.0088)
75	0.1765	0.1995	0.2135	0.3635	0.3960	0.3894	0.4538	0.4834	0.4595
	(0.0128)	(0.0136)	(0.0048)	(0.0134)	(0.0131)	(0.0061)	(0.0143)	(0.0171)	(0.0066)
100	0.1833	0.2003	0.2119	0.3729	0.3913	0.3923	0.4647	0.4777	0.4678
	(0.0097)	(0.0100)	(0.0038)	(0.0107)	(0.0090)	(0.0048)	(0.0110)	(0.0093)	(0.0046)

results of the simulation study showed that the bias-reduced YW estimator had the best performance in terms of bias compared with the YW and ML estimators and in terms of the MSE the ML estimators had the better results. Based on the results, we recommend the use of the reduced-bias YW as estimator of the α parameter of an NGINAR(1) process especially in small sample size case.

3.6 Appendix

Proof of Proposition 3.3.1

We have that the stationary NGINAR(1) process is α-mixing with exponentially decreasing weights, so the proof is similar to that of Schweer and Weiß (2014) Appendix A.4. there exist $a > 0$ and $0 < \nu < 1$ such that the sequence $(f_n)_{\mathbb{N}}$ of weights can be defined as $f_n = a \cdot \nu^n$. Consider the multivariate process ξ_t defined by

$$\xi_t := (X_t - \mu, X_t^2 - \mu(1+2\mu), X_t X_{t-1} - \mu(\mu + \alpha(1+\mu))).$$

Since ξ_t is a function of finitely many random variables for X_t, ξ_t is also α-mixing with exponentially decreasing weights $\tilde{f}_n = \tilde{a} \cdot \nu^n$. In particular, $\sum_{n\in\mathbb{N}} \tilde{f}_n^{1/2} < \infty$.

Applying a multivariate version of the result given on p.376 in Billingsley (1979), it follows that $T^{-1/2}\sum_{t=1}^{T} \xi_t$ has an asymptotically multivariate normal distribution with zero mean vector and covariance matrix $\Sigma = (\sigma_{i,j})$ with

$$\sigma_{i,j} = E(\xi_0^i \xi_0^j) + \sum_{k=1}^{\infty} E(\xi_0^i \xi_k^j) + \sum_{k=1}^{\infty} E(\xi_k^i \xi_0^j),$$

where ξ_t^h representing the h-th component of ξ_t. Using the expression for the joint moments given by Theorem (3.2.1), we have

$$\sigma_{1,1} = E[(\xi_0^1)^2] + 2\sum_{k=1}^{\infty} E[\xi_0^1 \xi_k^1] = \sigma_X^2 + 2\sum_{k=1}^{\infty} \gamma(k) = \sigma_X^2(1 + 2\sum_{k=1}^{\infty} \rho^k) = \sigma_X^2 \frac{1+\rho}{1-\rho}.$$

Now, we consider $\sigma_{1,2}$.

$$\sigma_{1,2} = E[\xi_0^1 \xi_0^2] + \sum_{k=1}^{\infty} E[\xi_0^1 \xi_k^2] + \sum_{k=1}^{\infty} E[\xi_k^1 \xi_0^2].$$

Using formulae of the joint moments, we obtain

$$\begin{aligned}
E[\xi_0^1\xi_0^2] &= \mu(0,0)-\mu\mu(0)=\mu(1+6\mu+6\mu^2)-\mu^2(1+2\mu)\\
&= (1+4\mu)\sigma_X^2,\\
E[\xi_0^1\xi_k^2] &= \mu(k,k)-\mu\mu(0)=2\mu\sigma_X^2\alpha^{2k}+2\mu\sigma_X^2\alpha^k+\frac{\alpha^k+\alpha^{k+1}-2\alpha^{2k+1}}{1-\alpha}\sigma_X^2\\
&\quad +\mu\mu(0)-\mu\mu(0)\\
&= 2\mu\sigma_X^2\alpha^{2k}+2\mu\sigma_X^2\alpha^k+\frac{\alpha^k+\alpha^{k+1}-2\alpha^{2k+1}}{1-\alpha}\sigma_X^2,\\
E[\xi_k^1\xi_0^2] &= \mu(0,k)-\mu\mu(0)=\mu(1+5\mu+4\mu^2)\alpha^k+\mu\mu(0)-\mu\mu(0)=(1+4\mu)\sigma_X^2\alpha^k.
\end{aligned}$$

Then,

$$\begin{aligned}
\sigma_{1,2} &= (1+4\mu)\sigma_X^2+2\mu\sigma_X^2\frac{\alpha^2}{1-\alpha^2}+2\mu\sigma_X^2\frac{\alpha}{1-\alpha}+\frac{\sigma_X^2}{1-\alpha}\left(\frac{\alpha}{1-\alpha}+\frac{\alpha^2}{1-\alpha}-\frac{2\alpha^3}{1-\alpha^2}\right)\\
&\quad +(1+4\mu)\sigma_X^2\frac{\alpha}{1-\alpha}\\
&= \frac{1+\alpha+\alpha^2-\alpha^3}{(1-\alpha)(1-\alpha^2)}\sigma^2+\frac{2+3\alpha+2\alpha^2}{1-\alpha^2}\mu\sigma_X^2.
\end{aligned}$$

For $\sigma_{1,3}$, we have

$$\sigma_{1,3}=\sum_{k=1}^{\infty}E[\xi_0^1\xi_k^3]+\sum_{k=0}^{\infty}E[\xi_k^1\xi_0^3].$$

So,

$$\begin{aligned}
E[\xi_0^1\xi_k^3] &= \mu(k-1,k)-\mu\mu(1)=\alpha\mu(k-1,k-1)+(1-\alpha)\mu\mu(k-1)-\mu\mu(1)\\
&= 2\mu\sigma_X^2\alpha^{2k-1}+2\mu\sigma_X^2\alpha^k+\frac{\alpha^k+\alpha^{k+1}-2\alpha^{2k}}{1-\alpha}\sigma_X^2+(1-\alpha)\mu\sigma_X^2\alpha^{k-1},\\
E[\xi_k^1\xi_0^3] &= \mu(1,k+1)-\mu\mu(1)=\alpha^k[\mu(1,1)-\mu\mu(1)].
\end{aligned}$$

Thus,

$$\begin{aligned}
\sigma_{1,3} &= 2\mu\sigma_X^2\frac{\alpha}{1-\alpha^2}+2\mu\sigma_X^2\frac{\alpha}{1-\alpha}+\left(\frac{\alpha}{1-\alpha}+\frac{\alpha^2}{1-\alpha}-\frac{2\alpha^2}{1-\alpha^2}\right)\frac{\sigma_X^2}{1-\alpha}+(1-\alpha)\frac{\mu\sigma_X^2}{1-\alpha}\\
&\quad +\frac{\mu(1,1)-\mu\mu(1)}{1-\alpha}.
\end{aligned}$$

We have

$$\begin{aligned}
\mu(1,1)-\mu\mu(1) &= (1+2\alpha)\alpha\sigma_X^2+2\alpha(1+\alpha)\mu\sigma_X^2+\mu\mu(0)-\mu\mu(1)\\
&= \alpha(1+2\alpha)\sigma_X^2+(\alpha+2\alpha^2+1)\mu\sigma_X^2.
\end{aligned}$$

Thus,

$$\sigma_{1,3} = \frac{(2+2\alpha-2\alpha^3)\alpha}{(1-\alpha)(1-\alpha^2)}\sigma_X^2+\frac{2+6\alpha+4\alpha^2+2\alpha^3}{1-\alpha^2}\mu\sigma_X^2.$$

For $\sigma_{2,2}$, we have

$$\sigma_{2,2} = E[(\xi_0^2)^2] + 2\sum_{k=1}^{\infty} E[\xi_0^2\xi_k^2].$$

Then,

$$E[(\xi_0^2)^2] \;=\; \mu(0,0,0) - (\mu(0))^2,$$

$$\begin{aligned}
E[\xi_0^2\xi_k^2] &= \mu(0,k,k) - \mu(0)^2 \\
&= \alpha^{2k}\mu(0,0,0) + \alpha(1+\alpha+2(1-\alpha)\mu)(\mu(1+5\mu+4\mu^2)\frac{\alpha^{k-1}-\alpha^{2k-1}}{1-\alpha} \\
&\quad +\mu\mu(0)\frac{1-\alpha^{2k}}{1-\alpha^2}) + \mu_{\epsilon,2}\mu(0)\frac{1-\alpha^{2k}}{1-\alpha^2} - \mu(0)^2 \\
&= \alpha^{2k}\mu(0,0,0) + \alpha(1+\alpha+2(1-\alpha)\mu)(\mu(1+5\mu+4\mu^2)\frac{\alpha^{k-1}-\alpha^{2k-1}}{1-\alpha} - \\
&\quad \mu\mu(0)\frac{\alpha^{2k}}{1-\alpha^2}) - \mu_{\epsilon,2}\mu(0)\frac{\alpha^{2k}}{1-\alpha^2} + \alpha(1+\alpha+2(1-\alpha)\mu)\frac{\mu\mu(0)}{1-\alpha^2} \\
&\quad + \frac{\mu_{\epsilon,2}\mu(0)}{1-\alpha^2} - \mu(0)^2.
\end{aligned}$$

where $\mu_{\epsilon,2} = \mathrm{E}[\epsilon_t^2] = \sigma_\epsilon^2 + \mu_\epsilon^2$.

We have

$$\alpha(1+\alpha+2(1-\alpha)\mu)\frac{\mu\mu(0)}{1-\alpha^2} + \frac{\mu_{\epsilon,2}\mu(0)}{1-\alpha^2} - \mu(0)^2 = 0.$$

So,

$$\begin{aligned}
E[\xi_0^2\xi_k^2] &= \alpha^{2k}\mu(0,0,0) + \alpha[1+\alpha+2(1-\alpha)\mu]\Big[\mu(1+5\mu+4\mu^2)\frac{\alpha^{k-1}-\alpha^{2k-1}}{1-\alpha} \\
&\quad -\mu\mu(0)\frac{\alpha^{2k}}{1-\alpha^2}\Big] - \mu_{\epsilon,2}\mu(0)\frac{\alpha^{2k}}{1-\alpha^2}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\sigma_{2,2} &= \mu(0,0,0) - (\mu(0))^2 + 2\sum_{k=1}^{\infty}\Big\{\alpha^{2k}\mu(0,0,0) - \alpha(1+\alpha+2(1-\alpha)\mu)\Big[\mu\mu(0)\frac{\alpha^{2k}}{1-\alpha^2} \\
&\quad -\mu(1+5\mu+4\mu^2)\frac{\alpha^{k-1}-\alpha^{2k-1}}{1-\alpha}\Big] - \mu_{\epsilon,2}\mu(0)\frac{\alpha^{2k}}{1-\alpha^2}\Big\}.
\end{aligned}$$

After a exhaustive algebra, and using the formula of the sum of the geometric series, we have

$$\sigma_{2,2} = \frac{1+\alpha+3\alpha^2-\alpha^3}{(1-\alpha)(1-\alpha^2)}\sigma_X^2 + \frac{11(1+\alpha^2)}{1-\alpha^2}\sigma_X^4 + \frac{1+11\alpha+5\alpha^2-\alpha^3}{(1-\alpha)(1-\alpha^2)}\mu\sigma_X^2 + \frac{9+16\alpha+9\alpha^2}{1-\alpha^2}\mu^2\sigma_X^2.$$

For $\sigma_{2,3}$, we have that,

$$\sigma_{2,3} = \sum_{k=1}^{\infty} E[\xi_0^2\xi_k^3] + \sum_{k=0}^{\infty} E[\xi_k^2\xi_0^3].$$

So,

$$\begin{aligned} E[\xi_0^2\xi_k^3] &= \mu(0,k-1,k) - \mu(0)\mu(1) \\ &= \mu(0,0,0)\alpha^{2k-1} + \alpha(1+\alpha+2(1-\alpha)\mu)\Big[(1+4\mu)\sigma_X^2(\frac{\alpha^{k-1}-\alpha^{2k-2}}{1-\alpha}) \\ &\quad - \mu\mu(0)\frac{\alpha^{2k-1}}{1-\alpha^2}\Big] - \mu_{\epsilon,2}\mu(0)\frac{\alpha^{2k-1}}{1-\alpha^2} + \mu(1-\alpha)(1+4\mu)\sigma_X^2\alpha^{k-1}, \end{aligned}$$

$$\begin{aligned} E[\xi_k^2\xi_0^3] &= \mu(1,k+1,k+1) - \mu(0)\mu(1) \\ &= \mu(1,1,1)\alpha^{2k} + \alpha(1+\alpha+2(1-\alpha)\mu)\Big[\mu(1,1)\frac{(\alpha^{k-1}-\alpha^{2k-1})}{1-\alpha} - \mu\mu(1)\Big(\frac{\alpha^{2k}}{1-\alpha^2} \\ &\quad + \frac{\alpha^{k-1}-\alpha^{2k-1}}{1-\alpha}\Big)\Big] - \mu_{\epsilon,2}\mu(1)\frac{\alpha^{2k}}{1-\alpha^2}. \end{aligned}$$

Thus,

$$\begin{aligned} \sigma_{2,3} &= \sum_{k=1}^{\infty}\Big\{\mu(0,0,0)\alpha^{2k-1} + \alpha(1+\alpha+2(1-\alpha)\mu)\Big[(1+4\mu)\sigma_X^2(\frac{\alpha^{k-1}-\alpha^{2k-2}}{1-\alpha}) \\ &\quad - \mu\mu(0)\frac{\alpha^{2k-1}}{1-\alpha^2}\Big] - \mu_{\epsilon,2}\mu(0)\frac{\alpha^{2k-1}}{1-\alpha^2} + \mu(1-\alpha)(1+4\mu)\sigma_X^2\alpha^{k-1}\Big\} \\ &\quad + \sum_{k=0}^{\infty}\Big\{\mu(1,1,1)\alpha^{2k} + \alpha(1+\alpha+2(1-\alpha)\mu)\Big[\mu(1,1)\frac{(\alpha^{k-1}-\alpha^{2k-1})}{1-\alpha} \\ &\quad - \mu\mu(1)\Big(\frac{\alpha^{2k}}{1-\alpha^2} + \frac{\alpha^{k-1}-\alpha^{2k-1}}{1-\alpha}\Big)\Big] - \mu_{\epsilon,2}\mu(1)\frac{\alpha^{2k}}{1-\alpha^2}\Big\}. \end{aligned}$$

Using the formula of the sum of the geometric series, $\sigma_{2,3}$ can be written as

$$\begin{aligned}\sigma_{2,3} &= \mu(0,0,0)\frac{\alpha}{1-\alpha^2} + \alpha(1+\alpha+2(1-\alpha)\mu)\Bigg[\frac{(1+4\mu)}{1-\alpha}\sigma_X^2(\frac{1}{1-\alpha}-\frac{1}{1-\alpha^2}) \\ &- \frac{\mu\mu(0)}{1-\alpha^2}\frac{\alpha}{1-\alpha^2}\Bigg] - \frac{\mu_{\epsilon,2}\mu(0)}{1-\alpha^2}\frac{\alpha}{1-\alpha^2} + \mu(1+4\mu)\sigma_X^2 + \frac{\mu(1,1,1)}{\alpha^2} \\ &+ \alpha(1+\alpha+2(1-\alpha)\mu)\Bigg\{\frac{\mu(1,1)}{1-\alpha}\left(\frac{\alpha^{-1}}{1-\alpha}-\frac{\alpha^{-1}}{1-\alpha^2}\right) \\ &- \mu\mu(1)\left[\frac{1}{(1-\alpha^2)^2}+\frac{1}{1-\alpha}\left(\frac{\alpha^{-1}}{1-\alpha}-\frac{\alpha^{-1}}{1-\alpha^2}\right)\right]\Bigg\} - \frac{\mu_{\epsilon,2}\mu(1)}{(1-\alpha^2)^2}.\end{aligned}$$

After a exhaustive and long algebra, we have

$$\begin{aligned}\sigma_{2,3} &= \frac{\alpha(2+21\alpha+4\alpha^2-11\alpha^3)}{(1-\alpha)(1-\alpha^2)}\mu\sigma_X^2 + \frac{(8+10\alpha+19\alpha^2-2\alpha^3)}{1-\alpha^2}\mu^2\sigma_X^2 \\ &+ \frac{(2+3\alpha+4\alpha^2-\alpha^3)}{(1-\alpha)(1-\alpha^2)}\alpha\sigma_X^2 + \frac{(6\alpha-3\alpha^2+12\alpha^3)}{1-\alpha^2}\sigma_X^4 + \frac{3\alpha^2(1+\alpha)\mu}{1-\alpha^2} \\ &+ \frac{(1+6\alpha^2-6\alpha^3)}{1-\alpha^2}\mu^2 - \frac{9\alpha^3}{1-\alpha^2}\mu^3 - \frac{1+3\alpha^2}{1-\alpha^2}\mu^4.\end{aligned}$$

Finally, we consider $\sigma_{3,3}$. We have that

$$\sigma_{3,3} = E[(\xi_0^3)^2] + 2\sum_{k=1}^{\infty} E[\xi_0^3\xi_k^3]. \tag{3.7}$$

So,

$$\begin{aligned}E[(\xi_0^3)^2] &= \mu(0,1,1) - [\mu(1)]^2 = \alpha^2\mu(0,0,0) + \alpha(1+\alpha+2(1-\alpha)\mu)\mu(0,0) \\ &+ \mu_{\epsilon,2}\mu(0) - \mu(1)^2,\end{aligned}$$

$$\begin{aligned}E[\xi_0^3\xi_k^3] &= \mu(1,k,k+1) - \mu(1)^2 = \alpha\mu(1,k,k) + (1-\alpha)\mu\mu(1,k) - \mu(1)^2 \\ &= \mu(1,1,1)\alpha^{2k-1} + \alpha(1+\alpha+2(1-\alpha)\mu)\left[\frac{\mu(1,1)}{1-\alpha}(\alpha^{k-1}-\alpha^{2k-2}) \right.\\ &\left. - \mu\mu(1)\left(\frac{\alpha^{2k-1}}{1-\alpha^2}+\frac{\alpha^{k-1}-\alpha^{2k-2}}{1-\alpha}\right)\right]\mu_{\epsilon,2}\mu(1)\frac{\alpha^{2k-1}}{1-\alpha^2} + (1-\alpha)\mu[\mu(1,1) \\ &- \mu\mu(1)]\alpha^{k-1}.\end{aligned}$$

Replacing this expressions in the equation (3.7) and using the formula of the sum of the geometric series, we have

$$\begin{aligned}\sigma_{3,3} &= \mu(1,1,1)\frac{2\alpha}{1-\alpha^2}+\alpha^2\mu(0,0,0)+\alpha(1+\alpha+2(1-\alpha)\mu)\mu(0,0)+\mu_{\epsilon,2}\mu(0)-\mu(1)^2\\ &+ [\mu(1,1)-\mu\mu(1)]\frac{2\alpha^2(1+\alpha+2(1-\alpha)\mu)}{(1-\alpha)(1-\alpha^2)}-\big[\alpha\mu(1+\alpha+2(1-\alpha)\mu)\\ &\quad+\mu_{\epsilon,2}\big]\frac{2\alpha\mu(1)}{(1-\alpha^2)^2}+[\mu(1,1)-\mu\mu(1)]\frac{2\mu}{1-\alpha}.\end{aligned}$$

Replacing $\alpha\mu(1+\alpha+2(1-\alpha)\mu)+\mu_{\epsilon,2}=(1-\alpha^2)\mu(0)$ and grouping the similar terms, we have

$$\begin{aligned}\sigma_{3,3} &= \mu(1,1,1)\frac{2\alpha}{1-\alpha^2}+\alpha^2\mu(0,0,0)+\alpha(1+\alpha+2(1-\alpha)\mu)\mu(0,0)+\mu_{\epsilon,2}\mu(0)\\ &- \big[\mu(1)(1-\alpha^2)+2\alpha\mu(0)\big]\frac{\mu(1)}{1-\alpha^2}+\big[2\mu(1+\alpha^2-2\alpha^3)\\ &+ 2\alpha^2(1+\alpha)\big]\frac{\mu(1,1)-\mu\mu(1)}{(1-\alpha)(1-\alpha^2)}. \end{aligned} \tag{3.8}$$

From Equation (3.3),

$$\begin{aligned}\mu(1,1,1) &= \alpha^3\mu(0,0,0)+3\alpha^2(1+\alpha+(1-\alpha)\mu)\mu(0,0)+\big[6\alpha\sigma_X^2-3\alpha^2\mu(0)\\ &+ \alpha(1+\alpha)(1+2\alpha)-9\alpha^3\mu\big]\mu(0)+6\mu^2\sigma_X^2(1-\alpha)-6\alpha^2\mu\sigma_X^2\\ &+ (1-\alpha-6\alpha^3)\mu^2. \end{aligned} \tag{3.9}$$

So, replacing (3.9) in (3.8) and grouping the similar terms, we have

$$\begin{aligned}\sigma_{3,3} &= \alpha^2\mu(0,0,0)\frac{(1+\alpha^2)}{1-\alpha^2}+\mu(0,0)\frac{\big[(1+\alpha)(\alpha+5\alpha^3)+(1-\alpha)(4\alpha^3+2\alpha)\mu\big]}{1-\alpha^2}\\ &\quad+\big[(1+10\alpha^2-6\alpha^3+\alpha^4)\sigma_X^2-(1-2\alpha-4\alpha^3-\alpha^4)\mu^2-(\alpha+\alpha^2-\alpha^3+17\alpha^4)\mu\\ &\quad+2\alpha^2(1+\alpha)(1+2\alpha)\big]\frac{\mu(0)}{1-\alpha^2}+\frac{10\alpha(1-\alpha)}{1-\alpha^2}\mu^2\sigma_X^2-\frac{12\alpha^3}{1-\alpha^2}+\frac{2\alpha(1-\alpha-6\alpha^3)}{1-\alpha^2}\mu^2\\ &\quad-\frac{\alpha^2(3-\alpha^2)}{1-\alpha^2}\sigma_X^4+\frac{1+2\alpha-\alpha^2}{1-\alpha^2}\mu^4+\big[2\mu(1+\alpha^2-2\alpha^3)\\ &\quad+2\alpha^2(1+\alpha)\big]\frac{\mu(1,1)-\mu\mu(1)}{(1-\alpha)(1-\alpha^2)}.\end{aligned}$$

Note that $\mu(0,0,0) = (1+12\sigma_X^2)\mu(0)$ and $\mu(0,0) = \mu + 6\mu\sigma_X^2$, so

$$\begin{aligned}
\sigma_{3,3} &= \big[\alpha^2(1+\alpha^2)(1+12\sigma_X^2) + (1+10\alpha^2-6\alpha^3+\alpha^4)\sigma_X^2 + (1-2\alpha-4\alpha^3-\alpha^4)\mu^2 \\
&- (\alpha+\alpha^2-\alpha^3+17\alpha^4)\mu + 2\alpha^2(1+\alpha)(1+2\alpha)\big]\frac{\mu(0)}{1-\alpha^2} + \big[(1+\alpha)(\alpha+5\alpha^3) \\
&+ (1-\alpha)(4\alpha^3+2\alpha)\mu\big]\frac{\mu}{1-\alpha^2} + \big[(1+\alpha)(\alpha+5\alpha^3) + (1-\alpha)(4\alpha^3+2\alpha)\mu\big]\frac{6\mu\sigma_x^2}{1-\alpha^2} \\
&+ 10\alpha(1-\alpha)\frac{\mu^2\sigma_X^2}{1-\alpha^2} - 12\alpha^3\frac{\mu\sigma_X^2}{1-\alpha^2} + 2\alpha(1-\alpha-6\alpha^3)\frac{\mu^2}{1-\alpha^2} - \alpha^2(3-\alpha^2)\frac{\sigma_X^4}{1-\alpha^2} \\
&+ (1+2\alpha-\alpha^2)\frac{\mu^4}{1-\alpha^2} + \big[\alpha(1+2\alpha)2\alpha^2(1+\alpha)\big]\frac{\sigma_X^2}{(1-\alpha)(1-\alpha^2)} \\
&+ \big[2(1+\alpha^2-2\alpha^3)(\alpha+2\alpha^2) + 2\alpha^2(1+\alpha)(1+\alpha+2\alpha^2)\big]\frac{\mu\sigma_X^2}{(1-\alpha)(1-\alpha^2)} \\
&+ \big[2(1+\alpha^2-2\alpha^3)(1+\alpha+2\alpha^2)\big]\frac{\mu^2\sigma_X^2}{(1-\alpha)(1-\alpha^2)}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\sigma_{3,3} &= \big[(1+22\alpha^2-6\alpha^3+13\alpha^4)\sigma_X^2 + (1-2\alpha-4\alpha^3-\alpha^4)\mu^2 - (\alpha+\alpha^2-\alpha^3+17\alpha^4)\mu \\
&+ 3\alpha^2+6\alpha^3+5\alpha^4\big]\frac{\mu(0)}{1-\alpha^2} + (8\alpha+6\alpha^2+18\alpha^3+18\alpha^4-34\alpha^5)\frac{\mu\sigma_X^2}{(1-\alpha)(1-\alpha^2)} \\
&+ (2+24\alpha-14\alpha^2+20\alpha^3-32\alpha^5)\frac{\mu^2\sigma_X^2}{(1-\alpha)(1-\alpha^2)} - \alpha^2(3-\alpha^2)\frac{\sigma_X^4}{1-\alpha^2} \\
&+ \big[2\alpha^3(1+\alpha)(1+2\alpha)\big]\frac{\sigma_X^2}{(1-\alpha)(1-\alpha^2)} + (4\alpha-4\alpha^2+4\alpha^3-16\alpha^4)\frac{\mu^2}{1-\alpha^2} \\
&+ (1+\alpha)(\alpha+5\alpha^3)\frac{\mu}{1-\alpha^2} + (1+2\alpha-\alpha^2)\frac{\mu^4}{1-\alpha^2}.
\end{aligned}$$

Finally, replacing $\mu(0) = \sigma_X^2 + \mu^2$ and after a tedious algebra, we have the following expression to $\sigma_{3,3}$:

$$\begin{aligned}
\sigma_{3,3} &= \frac{\alpha+2\alpha^2+10\alpha^3+5\alpha^4-6\alpha^5}{(1-\alpha)(1-\alpha^2)}\sigma_X^2 + \frac{1+19\alpha^2-6\alpha^3+14\alpha^4}{1-\alpha^2}\sigma_X^4 \\
&+ \frac{6\alpha+7\alpha^2+25\alpha^3-21\alpha^4-\alpha^5}{(1-\alpha)(1-\alpha^2)}\mu\sigma_X^2 + \frac{4+20\alpha+11\alpha^2-15\alpha^3+25\alpha^4-43\alpha^5}{(1-\alpha)(1-\alpha^2)}\mu^2\sigma_X^2 \\
&+ \frac{\alpha^2}{1-\alpha^2}\mu + \frac{4\alpha-\alpha^2}{1-\alpha^2}\mu^2 + \frac{2}{1-\alpha^2}\mu^4.
\end{aligned}$$

4 On shifted integer-valued autoregressive model for count time series showing equidispersion, underdispersion and overdispersion

Resumo

Neste capítulo, introduzimos o processo autorregressivo de primeira ordem com inovações Borel baseado no operador "thinning" binomial. Este modelo é adequado para séries temporais de contagem truncadas no zero que apresentam equidispersão, subdispersão e sobredispersão. Propriedades básicas do processo são obtidas. Para estimar os parâmetros desconhecidos são considerados os métodos de estimação de Yule-Walker, mínimos quadrados condicionais e máxima verossimilhança condicional. A distribuição assintótica dos estimadores de mínimos quadrados condicionais é obtida e testes de hipótese para um modelo equidisperso contra um subdisperso ou sobredisperso são formulados. Simulação de Monte Carlo é apresentada analisando a performance dos estimadores em amostras finitas. Duas aplicações com dados reais são apresentadas mostrando que o modelo Borel INAR(1) é adequado para dados de contagens com subdispersão e sobredispersão.

Palavras-chave: Distribuição de Borel. Normalidade assintótica. Operador thinning binomial. Séries temporais com valores inteiros.

Abstract

In this chapter, we introduce a first order integer-valued autoregressive process with Borel innovations based on the binomial thinning operator. This model is suitable to modelling zero truncated count time series with equidispersion, underdispersion and overdispersion. The basic properties of the process are obtained. To estimate the unknown parameters, the Yule-Walker (YW), conditional least squares (CLS) and conditional maximum likelihood (CML) methods are considered. The asymptotic distribution of CLS estimators is obtained and hypothesis tests to check an equidispersed model against an underdispersed or overdispersed model are formulated. A Monte Carlo simulation is presented in order to quantify the estimators performance in finite samples. Two applications to real data are presented to show that the Borel INAR(1) model is suited to model underdispersed and overdispersed data counts.

Keywords: Asymptotic normality. Binomial thinning operator. Borel distribution. Integer-valued time series model.

4.1 Introduction

Research in integer-valued time series has grown up in the last three decades. A reason for this is the numerous situations where we find time series with small integer values. The number of patients in a hospital at a specific point of time and the number of persons in a queue waiting for service at a certain moment (Alzaid and Al-Osh, 1987); the daily number of absent workers in a firm and the monthly number of ongoing strikes in a particular country, region or industry (Jung and Tremayne, 2006); the number of different IP address (Weiß , 2007) are examples of integer-valued data time series. The aim is to define a model for integer-valued time series that has covariance and properties similar structure to the continuous-valued autoregressive (AR) model.

McKenzie (1985) and Al-Osh and Alzaid (1987), independently, were the first to introduce an integer-valued autoregressive (INAR) process with Poisson marginals, Poisson INAR(1), based on the binomial thinning operator (Steutel and Van Harn, 1979). The equidispersion is the principal characteristic of the Poisson INAR(1) model, i.e., the mean is equal to variance, however many real-world data sets present overdispersion, variance larger than their mean, or underdispersion, variance smaller than their mean. For this reason, the Poisson INAR(1) model is not always suited for modelling integer-valued time series. In this sense, several models with overdispersion or underdispersion were defined in the recent years.

Ristić et al. (2009) defined the negative binomial thinning operator and based on the new thinning operator introduced the overdispersed integer-valued time series model with geometric marginal. Weiß (2013) investigated some distributions with underdispersion and combined these models with the INAR(1) process to obtain a process that presents underdispersion. Schweer and Weiß (2014) considered the compound Poisson INAR(1) processes (CPINAR(1)), which is suitable for modelling data sets with overdispersion. For such CPINAR(1) processes, explicit results were derived for joint moments, the k-step-ahead distribution, the stationary distribution and mixing properties. Closed-form expressions for the asymptotic distribution of the dispersion index for CPINAR(1) processes were derived and developed a test to overdispersion based on the index of dispersion. Bourguignon and Weiß (2017) defined a new thinning operator that extends the binomial and negative binomial thinning operators. Based on the new thinning introduced the BerG-INAR(1) process, a stationary count data with Bernoulli-geometric marginals. The BerG distribution was obtained as a convolution of the Bernoulli and geometric distributions. the BerG-INAR(1) process is suitable for time series data with equidispersion, underdispersion and overdispersion. They derived the probability generating function, moments, transition probabilities and zero probability. The maximum likelihood method was used for estimated the model parameters.

Although several first order non-negative integer valued time series models have been developed in recent years, there are some cases where are more suitable truncated models, for instance, integer-valued time series where zeros are not observed. The truncated models have not received much attention in research of count data times series, few works have been done in this case. Bakouch and Ristić (2010) introduced a new stationary INAR(1) process with zero truncated Poisson marginal distribution (ZTPINAR(1)). Several properties of the process were derived and applications to two crime data sets were presented. Nastić (2012) defined two types of shifted geometric INAR(1) models based on the negative binomial thinning operator. These models are suitable to positive count data and have a simple form with only two parameters. Their correlation properties are derived, conditional mean and conditional variance are considered. Statistical properties of the innovations process are obtained, non-parametric estimators of models parameters are defined and their asymptotic characterizations are given. Bourguignon and Vasconcellos (2015b) introduced a family of INAR(1) processes with power series innovations (PSINAR) which is suitable to modelling non-negative integer valued time series with equidispersion, overdispersion and underdispersion. The PSINAR(1) models has the flexibility of modelling positive integers data sets, this can be achieved by considering truncated distributions for the process innovations.

The purpose of this chapter is to introduce a shifted INAR(1) model with Borel innovations based on binomial thinning operator which exhibits equidispersion, overdispersion and underdispersion. The Borel INAR(1) (BINAR(1)) is a simple model with only two parameters suitable to zero truncated data sets. The process is an ergodic stationary Markov chain and its structure is different of other already proposed models. Although the BINAR(1) process in some cases shows equidispersion, in this cases, it does not coincide with the Poisson INAR(1) model.

The chapter is organized as follows. In Section 2, we introduce the Borel distribution proposed by Borel (1942), present its definition, properties and moments. Section 3 introduces the stationary Borel INAR model, its statistical properties are obtained, the transition probabilities, conditional mean and variance are derived. In Section 4, the parameter estimation is discussed; we consider CLS, YW and conditional maximum likelihood estimators. A simulation study is considered in Section 5. Two real data examples are presented in Section 6. Finally, we conclude the chapter in Section 7.

4.2 Borel distribution

In this section, we will briefly review about the Borel distribution (Borel, 1942). We present its definition and properties. Later in Section 4.3, we will consider the Borel distribution to introduce the Borel INAR(1) model with Borel distribution innovations which is adequate for truncated data set with equidispersion, overdispersion or underdispersion.

4.2.1 *Definition and properties*

The Borel distribution is a discrete probability distribution that was first obtained by Borel (1942). It is included in a subclass of a family of discrete probability distributions known as Lagrangian probability distributions and is classified within the so called basic Lagrangian distributions of the first kind. The Borel distribution is a special case of the queuing process and it can be viewed as a particular case of the Borel-Tanner distribution (Haight end Breuer, 1960). For more details of Lagrangian distributions, see Consul and Famoye (2006).

The probability mass function (pmf) from a discrete random variable Y with Borel distribution is given by

$$\mathrm{P}(Y = y) = \frac{(y\lambda)^{y-1}\mathrm{e}^{-\lambda y}}{y!}, \quad y = 1, 2, 3, \ldots, \tag{4.1}$$

where $0 < \lambda < 1$.

The Borel distribution has a simple interpretation in the theory of queues following Haight and Brauer (1960): “Suppose a queue is initiated with one member and let λ be equal to the traffic intensity assuming Poisson arrivals and constant service time. The pmf of Borel distribution, $\mathrm{P}(Y = y)$, represents the probability that exactly y members of queue will be served before the queue first vanishes.”

The properties of the Borel distribution can be found in Section 8.4 from Consul and Famoye (2006). In particular, the mean and variance of the distribution in (4.1) are given by

$$\mu_Y = (1-\lambda)^{-1} \quad \text{and} \quad \sigma_Y^2 = \lambda(1-\lambda)^{-3}.$$

The Borel distribution satisfies the properties of equidispersion, when $\lambda = 3/2 - \sqrt{5}/2$, underdispersion, when $0 < \lambda < 3/2 - \sqrt{5}/2$ and overdispersion, when $3/2 - \sqrt{5}/2 < \lambda < 1$.

We have that the probability generating function (pgf), $E(u^Y) := G_Y(u)$, and moment generating function (mgf), M_Y, of a random variable Y following the Borel distribution are, respectively, given by

$$G_Y(u) = z, \quad \text{wherezsatisfies} \quad z = u\,\mathrm{e}^{\lambda(z-1)},$$

and

$$M_Y(\beta) = \mathrm{e}^s, \quad \text{wheressatisfies} \quad s = \beta + \lambda(\mathrm{e}^s - 1).$$

All moments of the Borel distribution exist for $0 < \lambda < 1$. The kth noncentral moment, $\mu'_k = E(Y^k)$, and the central moment, $\mu_k = E(Y - \mu_Y)^k$, satisfy the following recurrence relations:

$$\mu'_{k+1} = \lambda(1-\lambda)^{-1}\frac{d\mu'_k}{d\lambda} + \mu'_1\mu'_k, \quad k = 0, 1, 2, \ldots,$$

and

$$\mu_{k+1} = \lambda(1-\lambda)^{-1}\frac{d\mu_k}{d\lambda} + k\mu_2\mu_{k-1}, \quad k = 1, 2, 3, \ldots,$$

where $\mu'_1 = (1-\lambda)^{-1}$ and $\mu_2 = \sigma_Y^2$.

In particular, the third and fourth central moments are given by

$$\mu_3 = \lambda(1+2\lambda)(1-\lambda)^{-5} \quad \text{and} \quad \mu_4 = \lambda(1+8\lambda+6\lambda^2)(1-\lambda)^{-7} + 3\mu_2^2. \tag{4.2}$$

Observe that, in the particular case that one the Borel distribution is equidispersed, where $\lambda = \lambda_0 = 3/2 - \sqrt{5}/2$, we have $\mu_Y = \sigma_Y^2 = 1/(1-\lambda_0)$ but different of μ_3 given in (4.2). Furthermore, the Poisson distribution with parameter $1/(1-\lambda_0)$ has mean, variance and μ_3 equal to $1/(1-\lambda_0)$ being different of the equidispersed Borel distribution.

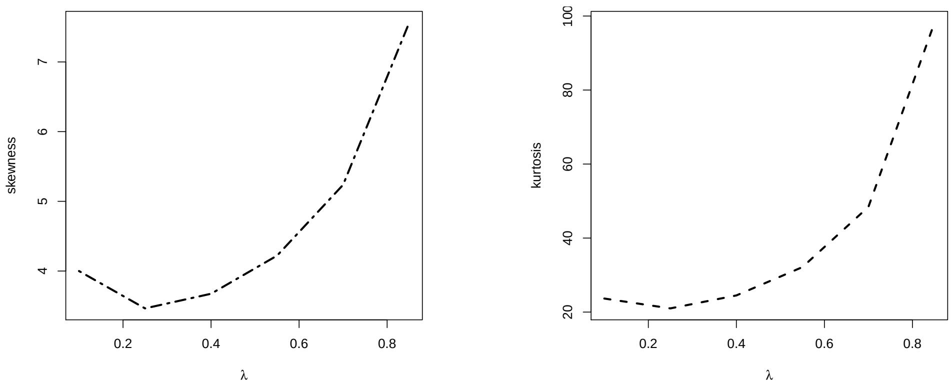
The coefficients of skewness, β_1, and kurtosis, β_2, are given by

$$\beta_1 = \frac{1+2\lambda}{\sqrt{\lambda(1-\lambda)}} \quad \text{and} \quad \beta_2 = 3 + \frac{1+8\lambda+6\lambda^2}{\lambda(1-\lambda)}. \tag{4.3}$$

Figure 4.1 displays the skewness and kurtosis of the Borel distribution. From 4.3 and Figure (4.1), we have that the Borel distribution is always positively skewed and leptokurtic.

4.3 Borel INAR(1) model

Let $\mathbb{Z}$ be the set of all integers, $\mathbb{N}$ be the set of positive integers and $\mathbb{N}_0$ the set of non-negative integers. In this section, we will introduce the first order integer-valued autoregressive process with Borel distribution innovations on $\mathbb{N}$, Borel INAR(1). This model is suitable for positive counting processes with equidispersion, overdispersion and underdispersion. Basic properties of the process are derived. The autocorrelation function, transition probabilities, conditional mean and variance are obtained.



(a) Skewness (b) Kurtosis

Figure 4.1: Skewness and kurtosis of the Borel distribution

4.3.1 *Definition and properties of the process*

Let X be a non-negative integer valued random variable and $\alpha \in [0, 1]$. Steutel and Van Harn (1979) introduced the binomial thinning operator as

$$\alpha \circ X = \sum_{i=1}^{X} W_i, \tag{4.4}$$

where W_i are independent and identically distributed binary random variables, independent of X with $\Pr(W_i = 1) = 1 - \Pr(W_i = 0) = \alpha$, i.e., W_i has Bernoulli distribution with α parameter. The sequence $\{W_i\}_{i\in\mathbb{N}}$ is called the counting series for $\alpha \circ X$. The properties of the binomial thinning operator can be found in Silva and Oliveira (2004) and Silva (2005). Based on this operator, we define the first-order integer-valued autoregressive process with Borel innovations.

Definition 4.3.1. *(Borel INAR(1))* The first-order integer-valued autoregressive process with Borel innovations $\{X_t\}_{t\in\mathbb{Z}}$, BINAR(1), is defined by equation

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{4.5}$$

where $\alpha \in [0, 1)$, $\alpha \circ X_{t-1}$ is the binomial thinning operator defined in (4.4), $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is the sequence of innovations with Borel(λ) distribution and ϵ_t is independent of W_{t-i} for all $i \geq 1$. Since that $\alpha \in [0, 1)$ and $\mu_\epsilon, \sigma^2_\epsilon < \infty$, from Section 3 and Theorem 2.1 by Du and Li (1991), the $\{X_t\}_{t\in\mathbb{Z}}$ process in (4.5) is an ergodic stationary Markov chain. The transition probabilities are given by

$$\Pr(X_t = k|X_{t-1} = l) = \sum_{i=0}^{\min(k-1,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \frac{[(k-i)\lambda]^{k-i-1}}{(k-i)!} \mathrm{e}^{-\lambda(k-i)}. \tag{4.6}$$

It is easy to see that, if $\mu_\epsilon, \sigma_\epsilon^2 < \infty$, the mean and variance of the INAR(1) process are given by

$$\mu_X = \mu_\epsilon/(1-\alpha) \quad \text{and} \quad \sigma_X^2 = (\sigma_\epsilon^2 + \alpha\mu_\epsilon)/(1-\alpha^2),$$

where μ_ϵ and σ_ϵ^2 are the mean and variance of innovations, respectively. Thus, in the BINAR(1) process, we have

$$\mu_X = \mathrm{E}(X_t) = \frac{1}{(1-\alpha)(1-\lambda)} \quad \text{and} \quad \sigma_X^2 = \mathrm{Var}(X_t) = \frac{\alpha(1-\lambda)^2 + \lambda}{(1-\alpha^2)(1-\lambda)^3}.$$

The dispersion index of X_t in INAR(1) model, I_X, is directly linked with the dispersion index of the innovations, $I_\epsilon = \sigma_\epsilon^2/\mu_\epsilon$, by the expression

$$I_X = \frac{\sigma_X^2}{\mu_X} = \left(1 + \frac{I_\epsilon}{\alpha}\right) \cdot \left(1 + \frac{1}{\alpha}\right)^{-1}.$$

Thus, we have that the dispersion of the marginal distribution depends of the innovations dispersion. The BINAR(1) dispersion index is given by

$$I_X = \frac{\lambda/(1-\lambda)^2 + \alpha}{1+\alpha}.$$

Since we have Borel innovations, it follows that this model presents

- equidispersion when $\lambda = 3/2 - \sqrt{5}/2$,
- overdispersion when $3/2 - \sqrt{5}/2 < \lambda < 1$,
- underdispersion when $0 < \lambda < 3/2 - \sqrt{5}/2$.

We have that the Borel INAR(1) model is suited when the time series data present overdispersion or underdispersion.

The conditional mean and variance of X_t given X_{t-1} are, respectively

$$\mathrm{E}(X_t|X_{t-1}) = \alpha X_{t-1} + \frac{1}{(1-\lambda)}$$

and

$$\mathrm{Var}(X_t|X_{t-1}) = \alpha(1-\alpha)X_{t-1} + \frac{3\,\lambda - 1 - \lambda^2}{(1-\lambda)^3} + \frac{1}{(1-\lambda)}.$$

It is also easy to verify that the autocorrelation function (ACF) at lag k is given by

$$\mathrm{Corr}(X_t, X_{t-k}) = \rho(k) = \alpha^k,\ k \geq 1, \tag{4.7}$$

which clearly is restricted to be positive and have the same behaviour as that of the AR(1) model.

4.4 Estimation and inference of the unknown parameters

In practice, the true values of the model parameters α and λ are not known but have to be estimated from a given time series data. This section is concerned with the estimation of the two parameters of interest. We consider three estimation methods, namely, CLS, YW, ML and CML.

4.4.1 *Conditional least squares estimation*

Let $X_1, X_2, \ldots, X_T$ stem from a stationary BINAR(1) process $\{X_t\}_{t\in\mathbb{Z}}$. The CLS estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\lambda})^{\mathsf{T}}$ of $\boldsymbol{\theta} = (\alpha, \lambda)^{\mathsf{T}}$ is given by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}(Q_T(\boldsymbol{\theta})),$$

where $Q_T(\boldsymbol{\theta}) = \sum_{t=2}^{T}[X_t - \mathrm{E}(X_t|X_{t-1})]^2$. Thus, following Klimko and Nelson (1978), the CLS estimators of α and λ can be written in closed form as

$$\widehat{\alpha}_{\text{CLS}} = \frac{\sum\limits_{t=2}^{T} X_t X_{t-1} - \dfrac{1}{T-1}\sum\limits_{t=2}^{T} X_t \sum\limits_{t=2}^{T} X_{t-1}}{\sum\limits_{t=2}^{T} X_{t-1}^2 - \dfrac{1}{T-1}\left(\sum\limits_{t=2}^{T} X_{t-1}\right)^2} \tag{4.8}$$

and

$$\widehat{\lambda}_{\text{CLS}} = 1 - \frac{T-1}{\left(\sum\limits_{t=2}^{T} X_t - \widehat{\alpha}_{\text{CLS}} \sum\limits_{t=2}^{T} X_{t-1}\right)},$$

where $\widehat{\alpha}_{\text{CLS}}$ is given in (4.8).

We establish the asymptotic normality of CLS estimators in the following Theorem having proof given in the Appendix.

Theorem 4.4.1. *The CLS estimators $\widehat{\alpha}_{\text{CLS}}$ and $\widehat{\lambda}_{\text{CLS}}$ are strongly consistent and asymptotic normal, i.e.*

$$T^{1/2}(\widehat{\alpha}_{\text{CLS}} - \alpha, \widehat{\lambda}_{\text{CLS}} - \lambda) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$$

with

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{\alpha}^2 & \sigma_{\alpha\lambda}^2 \\ \sigma_{\alpha\lambda}^2 & \sigma_{\lambda}^2 \end{pmatrix},$$

where

$$
\begin{aligned}
\sigma_{\alpha}^{2} &= (1-\alpha^{2}) + \frac{\alpha(1-\alpha)(1-\alpha^{2})^{2}(1-\lambda)^{5}}{(1-\alpha^{3})(\lambda+\alpha(1-\lambda)^{2})^{2}}\left[\frac{\alpha}{1-\alpha} + \frac{3\alpha^{3}}{1+\alpha} + \frac{3\alpha^{2}\lambda}{(1+\alpha)(1-\lambda)^{2}} + \frac{1+2\lambda}{(1-\lambda)^{4}}\right], \\
\sigma_{\alpha\lambda}^{2} &= \alpha(1-\alpha)(1-\lambda)^{2} - \frac{\alpha(1+\alpha)(1-\lambda)^{3}}{\lambda+\alpha(1-\lambda)^{2}} - [(1+\alpha)+3\alpha^{2}(1-\alpha)]\frac{\alpha^{2}(1-\alpha^{2})(1-\lambda)^{7}}{[\lambda+\alpha(1-\lambda)^{2}]^{2}} \\
&\quad - \frac{\lambda(1-\alpha)(1-\lambda)}{\lambda+\alpha(1-\lambda)^{2}} - [3\alpha^{2}\lambda(1-\lambda)^{2} + (1+2\lambda)(1+\alpha)]\frac{\alpha(1-\alpha)^{2}(1+\alpha)(1-\lambda)^{3}}{[\lambda+\alpha(1-\lambda)^{2}]^{2}}, \\
\sigma_{\lambda}^{2} &= \{1-\alpha[\alpha\lambda+(1-2\alpha)(1+\lambda^{2})+\alpha^{2}(1-\lambda)^{2}]\}\frac{1-\lambda}{(1-\alpha)^{2}} + \frac{\alpha(1+\alpha)(1-\lambda)^{4}}{(1-\alpha)[\lambda+\alpha(1-\lambda)^{2}]} \\
&\quad + \frac{3\alpha^{3}(1-\lambda)^{5}(1-\alpha^{2})}{(1-\alpha^{3})[\lambda+\alpha(1-\lambda)^{2}]} + [(1-\alpha)(1+2\lambda)+\alpha(1-\lambda)^{4}]\frac{\alpha(1+\alpha)^{2}(1-\lambda)^{3}}{(1-\alpha^{3})[\lambda+\alpha(1-\lambda)^{2}]^{2}}.
\end{aligned}
$$

Considering $\lambda_0 = 3/2 - \sqrt{5}/2$, the asymptotic normality of $\widehat{\lambda}_{\text{CLS}}$ is particularly useful if our interest is to test the null hypothesis of an equidispersed model, $H_0 : \lambda = \lambda_0$, against the alternative hypothesis of an underdispersed model $H_1 : \lambda < \lambda_0$ or an overdispersed model, $H_1 : \lambda > \lambda_0$.

Based on the asymptotic normality of $\widehat{\lambda}_{CLS}$ estimator, we reject $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda < \lambda_0$ on significance level δ if

$$\widehat{\lambda}_{CLS} < \lambda_0 + q_\delta \times \sqrt{\widehat{\sigma}^2_{\lambda_0}/T} \tag{4.9}$$

and we reject $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda > \lambda_0$ on significance level δ if

$$\widehat{\lambda}_{CLS} > \lambda_0 + q_{1-\delta} \times \sqrt{\widehat{\sigma}^2_{\lambda_0}/T}, \tag{4.10}$$

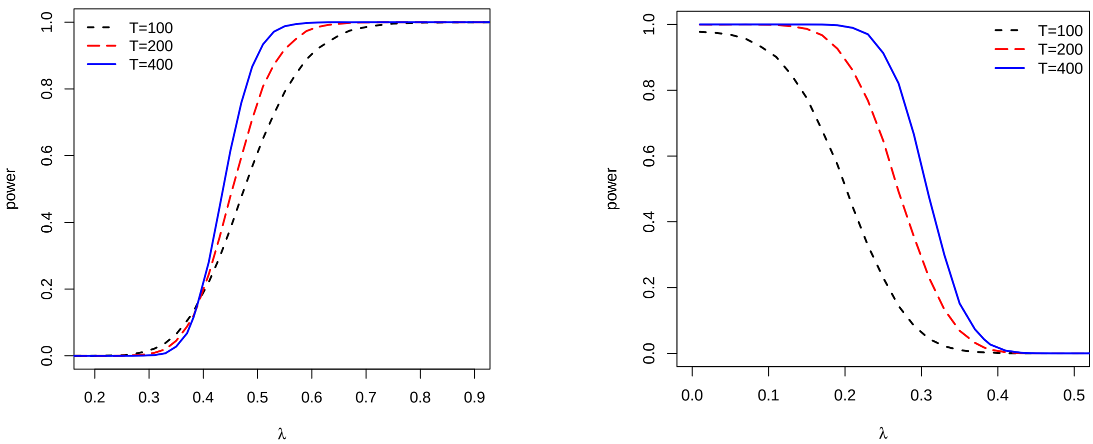
where $\widehat{\sigma}^2_{\lambda_0}$ is the variance of $\widehat{\lambda_{CLS}}$ obtained replacing λ_0 in the σ^2_λ expression, q_δ is the δ-quantile of the standard normal distribution, that is, $\Phi(q_\delta) = \delta$, $\delta \in (0, 1)$ and Φ is the standard normal cumulative distribution function. Alternatively, we will reject $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda < \lambda_0$ (overdispersion) if

$$\Phi\left((\widehat{\lambda} - \lambda_0)/\sqrt{\widehat{\sigma}^2_{\lambda_0}/T}\right) < \delta.$$

If the alternative hypothesis of interest is $H_1 : \lambda > \lambda_0$ (underdispersion), we reject $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda > \lambda_0$ when

$$1 - \Phi\left((\widehat{\lambda} - \lambda_0)/\sqrt{\widehat{\sigma}^2_{\lambda_0}/T}\right) < \delta.$$

Figure 4.2 shows the behaviour of the power functions against λ for sample sizes $T = 100, 200, 400$. We consider the significance level $\delta = 0.05$, $\alpha = 0.15$ and $\lambda \in (0, 1)$. We can



(a) $H_1 : \lambda > \lambda_0$ (b) $H_1 : \lambda < \lambda_0$

Figure 4.2: Power function against λ for BINAR(1) model.

see, as expected, that the power of dispersion test becomes better as we increase the sample size T.

Tables 4.1 and 4.2 present results from a simulation to access the quality of asymptotic approximations in Theorem 4.4.1. We simulated the underdispersed Borel INAR(1) process with $(\alpha, \lambda) = (0.2, 0.2)$ and the overdispersed Borel INAR(1) process with $(\alpha, \lambda) = (0.2, 0.5)$, respectively. We considered the significance level $\delta = 0.05$ and 10000 replications in both cases. In the same way of Schweer and Weiß (2014), we show both the empirically observed results and the values obtained from the asymptotic approximation.

Table 4.1: Simulated Borel INAR(1) with $(\alpha, \lambda) = (0.2, 0.2)$, significance level $\delta = 0.05$ and 10000 replications: Descriptive statistics and asymptotic properties of $\widehat{\lambda}$; Empirical and asymptotic rejection rates (r.r.).

T	Mean	s.d.	s.d.$_a$	Skewness	$\widehat{q}_{0.05}$	$q_{0.05,a}$	$\widehat{q}_{0.95}$	$q_{0.95,a}$	r.r.$_\alpha$	r.r.$_{\widehat{\alpha}}$	r.r.$_a$
100	0.2012	0.1168	0.1066	-0.9253	-0.0100	0.0245	0.3639	0.3754	0.4688	0.4764	0.4660
250	0.2003	0.0731	0.0674	-0.5094	0.0672	0.0890	0.3085	0.3109	0.8597	0.8894	0.8375
500	0.2021	0.0524	0.0476	-0.3405	0.1114	0.1215	0.2833	0.2784	0.9874	0.9927	0.9837

The null rejection rates of a Borel INAR(1) with parameters (α, λ) are presented in both tables, where the critical value according to (4.9) in the underdispersed case and according to (4.10) in the overdispersed case was computed either with the true α (r.r.$_\alpha$) or by plugging-in CLS estimator $\widehat{\alpha}$ (r.r.$_{\widehat{\alpha}}$). We observe, in both cases, that the rejection rate increases as we increase the sample size, to both for α and $\widehat{\alpha}$ and the asymptotic approximating $(\mathrm{r.r.}_a)$ and the empirical values of reject rates are be very close to the ones in the asymptotic approximate. In general, the empirical results of the quantiles $\widehat{\mathrm{q}}_\beta$ with $\beta = 0.05, 0.95$ and the standard deviation are very close to its asymptotic results for $T \geq 250$ as expected. We conclude that the asymptotic results have reasonable approximations for finite samples in both, underdispersed and overdispersed cases.

Table 4.2: Simulated Borel INAR(1) with $(\alpha, \lambda) = (0.2, 0.5)$, significance level $\delta = 0.05$ and 10.000 replicas: Descritive statistics and asymptotic properties of $\widehat{\lambda}$; empirical and asymptotic rejection rates.

T	Mean	s.d.	s.d.$_a$	Skewness	$\widehat{\mathrm{q}}_{0.05}$	$\mathrm{q}_{0.05,a}$	$\widehat{\mathrm{q}}_{0.95}$	$\mathrm{q}_{0.95,a}$	r.r.$_\alpha$	r.r.$_{\widehat{\alpha}}$	r.r.$_a$
100	0.4970	0.0840	0.0821	-1.0474	0.3421	0.3648	0.6351	0.6100	0.4647	0.4781	0.4527
250	0.4980	0.0522	0.0519	-0.6039	0.4052	0.4145	0.5753	0.5854	0.7303	0.7189	0.7469
500	0.4998	0.0376	0.0367	-0.4280	0.4343	0.4395	0.5569	0.5604	0.9150	0.9003	0.9272

4.4.2 *Yule-Walker estimation*

Let $X_1, \ldots, X_T$ stem from a stationary BINAR(1) process $\{X_t\}_{t\in\mathbb{Z}}$. The sample autocorrelation function is given by

$$\widehat{\rho}(k) = \frac{\sum_{t=1}^{T-k}(X_t - \overline{X})(X_{t+k} - \overline{Y})}{\sum_{t=1}^{n}(X_t - \overline{X})^2},$$

where $\overline{X} = (1/T)\sum_{t=1}^{T} X_t$ is the sample mean. The Yule-Walker (YW) estimator of α, based upon the fact that $\rho(k) = \alpha^k$, as in (4.7), is given by

$$\widehat{\alpha}_{\mathrm{YW}} = \widehat{\rho}(1) = \frac{\sum_{t=1}^{T-1}(X_t - \overline{X})(X_{t+1} - \overline{X})}{\sum_{t=1}^{T}(X_t - \overline{X})^2}. \tag{4.11}$$

The first moment of $\{X_t\}_{t\in\mathbb{Z}}$ is given by $\mathrm{E}(X_t) = 1/(1-\alpha)(1-\lambda)$. Using this, the estimator of λ is defined as

$$\widehat{\lambda}_{\mathrm{YW}} = 1 - \frac{1}{\overline{X}(1 - \widehat{\alpha}_{\mathrm{YW}})},$$

where $\widehat{\alpha}_{\mathrm{YW}}$ is given in (4.11).

Freeland and McCabe (2005) considering a Poisson INAR(1) process showed that the YW and CLS estimators are asymptotically equivalent. The next proposition, which proof is in the Appendix, shows that the result also holds for the BINAR(1) process.

Theorem 4.4.2. *In the BINAR(1) process*

$$\begin{pmatrix} \widehat{\alpha}_{\text{CLS}} - \widehat{\alpha}_{\text{YW}} \\ \widehat{\lambda}_{\text{CLS}} - \widehat{\lambda}_{\text{YW}} \end{pmatrix} = o_p(T^{-1/2}),$$

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

In section 4.5 we compare CLS and YW estimators and we see that the results of both are similar as expected.

4.4.3 *Conditional maximum likelihood estimation*

Suppose that X_1 is fixed. The conditional log-likelihood function is then given by

$$\ell(\boldsymbol{\theta}) = \log\left[\prod_{t=2}^{T} \Pr(X_t = k|X_{t-1} = l)\right] = \sum_{t=2}^{T} \log[\Pr(X_t = k|X_{t-1} = l)],$$

with $\Pr(X_t = k|X_{t-1} = l)$ as in (4.6).

The CML estimator of $\boldsymbol{\theta}$ is the value $\widehat{\boldsymbol{\theta}}_{\text{CML}} = (\widehat{\alpha}_{\text{CML}}, \widehat{\lambda}_{\text{CML}})^{\mathsf{T}}$ that maximizes $\ell(\boldsymbol{\theta})$. Since $\ell'(\boldsymbol{\theta})$ is a nonlinear function, the maximum likelihood estimate of $\boldsymbol{\theta}$ must be computed by using numerical methods.

CML estimators, as noted above, do not have closed-form expressions and Fisher's information matrix, $\mathbf{K}(\boldsymbol{\theta})$, is also not available. In practice, the observed Fisher information $(T-1)\mathbf{J}(\boldsymbol{\theta})$ can be used to approximate $\mathbf{K}(\boldsymbol{\theta})$, and plugging in the obtained estimates allows, for instance, to approximate the asymptotic standard errors of the CML estimators; also see Freeland and McCabe (2004).

4.5 Estimators numerical comparison

In this section, we report the results of Monte Carlo simulation experiments that were performed to evaluate the behaviour of the CLS, YW and CML estimators. The Monte Carlo simulation experiments were performed using the R programming language; see http://www.r-project.org. The CML estimates of α and λ were obtained by maximizing the conditional log-likelihood function using the BFGS quasi-Newton nonlinear optimization algorithm with numerical derivatives. As initial values for the algorithm, we considered the estimates obtained by CLS method. We performed 5000 replications and the sample sizes considered were T = 50, 100, 200 and 400. For the values of parameters, we considered $\alpha = 0.15, 0.3, 0.45, 0.6, 0.75$ and $\lambda = 0.3, 0.6$. Tables 4.3 and 4.4 present the empirical means and mean squared errors of the estimates of the parameters of the BINAR(1) process. From the results we conclude that CML estimates have smaller bias, as expected since it uses the whole information of the distribution, and the bias tends to zero for all estimators as the sample size increases, since that the estimators are asymptotically unbiased. Although the CML has the best performance, the CLS estimator is more indicated because it has closed form and its performance is very close to that of CML estimator.

4.6 Empirical illustrations

Here, we present two applications of the Borel INAR(1) model to a real data set. First, in Section 4.6.1, we consider an underdispersed real data set and second, in Subsection 4.6.2, we fitted an overdispersed real data set to BINAR(1) model. We considered the CML method to estimate the unknown parameters of the models in both cases. After, we considered the $\widehat{\lambda}_{\mathrm{CLS}}$ estimator and the asymptotic distribution given in Theorem 4.4.1. So, we carried out a hypothesis test in both applications to test the hypothesis $H_0 : \lambda = 3/2 - \sqrt{5}/2$ against $H_1 : \lambda < 3/2 - \sqrt{5}/2$ (underdispersed model) and $H_0 : \lambda = 3/2 - \sqrt{5}/2$ against $H_1 : \lambda > 3/2 - \sqrt{5}/2$ (overdispersed model). The required numerical evaluations were carried out using the R programming language as in Section 4.5.

4.6.1 *Modeling underdispersion*

Now, we present an application of the Borel INAR(1) model to the real data set in the underdispersed case. The data set is given by Bakouch and Ristić (2010) as an application of the ZTPINAR(1) process. Their original data are counts of vagrancy in the 64th police car beat in Pittsburgh, during one month. The data consist of 144 observations, starting in January 1990 and ending in December 2001 and can be obtained from the crime data section of the forecasting principles site (http://www.forecastingprinciples.com). With the aim to use the ZTPINAR(1) model, the authors transformed the series adding 1 to each observation. Here we also consider the transformed series.

In Table 4.5, we present some empirical descriptive measures for the data set of counts of vagrancy, which include central tendency statistics, variance, skewness and kurtosis, among others. We observe that the empirical mean is much larger than the empirical variance, showing the data are underdispersed.

The time series data and their sample autocorrelation and partial autocorrelation are displayed in Figure 4.3. From Figure 4.3, we see that a first order autoregressive model may be appropriate for the given data series because of the clear cut-off after lag 1 in the partial autocorrelations.

Table 4.3: Empirical means and mean squared errors (in parentheses) of the estimates of the parameters for $\lambda = 0.3$ and some values of α and T.

		Estimator of α			Estimator of λ		
T	α	$\widehat{\alpha}_{\text{YW}}$	$\widehat{\alpha}_{\text{CLS}}$	$\widehat{\alpha}_{\text{CML}}$	$\widehat{\lambda}_{\text{YW}}$	$\widehat{\lambda}_{\text{CLS}}$	$\widehat{\lambda}_{\text{CML}}$
50	0.15	0.1171	0.1196	0.1554	0.3002	0.2968	0.2833
		(0.0210)	(0.0217)	(0.0050)	(0.0186)	(0.0197)	(0.0071)
	0.30	0.2530	0.2588	0.3024	0.3101	0.3029	0.2832
		(0.0226)	(0.0230)	(0.0046)	(0.0234)	(0.0259)	(0.0076)
	0.45	0.3884	0.3968	0.4489	0.3303	0.3186	0.2859
		(0.0218)	(0.0214)	(0.0037)	(0.0280)	(0.0301)	(0.0073)
	0.60	0.5277	0.5396	0.5996	0.3556	0.3345	0.2859
		(0.0212)	(0.0201)	(0.0023)	(0.0387)	(0.0431)	(0.0073)
	0.75	0.6633	0.6797	0.7485	0.4150	0.3743	0.2895
		(0.0199)	(0.0175)	(0.0010)	(0.0565)	(0.0677)	(0.0070)
100	0.15	0.1319	0.1332	0.1532	0.3017	0.3005	0.2931
		(0.0115)	(0.0116)	(0.0025)	(0.0102)	(0.0104)	(0.0034)
	0.30	0.2746	0.2776	0.3016	0.3069	0.3037	0.2923
		(0.0116)	(0.0116)	(0.0023)	(0.0121)	(0.0126)	(0.0035)
	0.45	0.4192	0.4237	0.4506	0.3147	0.3088	0.2935
		(0.0107)	(0.0106)	(0.0017)	(0.0164)	(0.0171)	(0.0034)
	0.60	0.5643	0.5702	0.5994	0.3261	0.3156	0.2933
		(0.0091)	(0.0087)	(0.0011)	(0.0223)	(0.0235)	(0.0034)
	0.75	0.7066	0.7143	0.7486	0.3585	0.3383	0.2948
		(0.0077)	(0.0071)	(0.0005)	(0.0333)	(0.0355)	(0.0035)
200	0.15	0.1415	0.1421	0.1513	0.2995	0.2989	0.2960
		(0.0060)	(0.0060)	(0.0012)	(0.0050)	(0.0051)	(0.0017)
	0.30	0.2872	0.2887	0.3000	0.3030	0.3015	0.2972
		(0.0060)	(0.0061)	(0.0011)	(0.0066)	(0.0068)	(0.0017)
	0.45	0.4349	0.4371	0.4497	0.3064	0.3034	0.2968
		(0.0052)	(0.0052)	(0.0009)	(0.0081)	(0.0083)	(0.0018)
	0.60	0.5825	0.5855	0.5995	0.3121	0.3070	0.2972
		(0.0043)	(0.0042)	(0.0006)	(0.0117)	(0.0120)	(0.0017)
	0.75	0.7289	0.7328	0.7495	0.3281	0.3174	0.2962
		(0.0031)	(0.0029)	(0.0003)	(0.0180)	(0.0188)	(0.0018)
400	0.15	0.1456	0.1460	0.1506	0.3001	0.2998	0.2984
		(0.0030)	(0.0030)	(0.0006)	(0.0026)	(0.0026)	(0.0008)
	0.30	0.2927	0.2934	0.2999	0.3022	0.3015	0.2985
		(0.0031)	(0.0031)	(0.0005)	(0.0033)	(0.0033)	(0.0008)
	0.45	0.4413	0.4424	0.4496	0.3048	0.3034	0.2990
		(0.0026)	(0.0026)	(0.0004)	(0.0041)	(0.0041)	(0.0008)
	0.60	0.5912	0.5927	0.6004	0.3072	0.3046	0.2983
		(0.0020)	(0.0019)	(0.0003)	(0.0058)	(0.0059)	(0.0009)
	0.75	0.7396	0.7416	0.7500	0.3149	0.3095	0.2980
		(0.0013)	(0.0013)	(0.0001)	(0.0090)	(0.0092)	(0.0009)

Table 4.4: Empirical means and mean squared errors (in parentheses) of the estimates of the parameters for $\lambda = 0.6$ and some values of α and T.

		Estimator of α			Estimator of λ		
T	α	$\widehat{\alpha}_{\text{YW}}$	$\widehat{\alpha}_{\text{CLS}}$	$\widehat{\alpha}_{\text{CML}}$	$\widehat{\lambda}_{\text{YW}}$	$\widehat{\lambda}_{\text{CLS}}$	$\widehat{\lambda}_{\text{CML}}$
50	0.15	0.1186	0.1210	0.1553	0.5926	0.5902	0.5826
		(0.0173)	(0.0182)	(0.0031)	(0.0095)	(0.0103)	(0.0060)
	0.30	0.2588	0.2645	0.3049	0.5975	0.5930	0.5826
		(0.0186)	(0.0190)	(0.0031)	(0.0113)	(0.0124)	(0.0061)
	0.45	0.3971	0.4059	0.4514	0.6075	0.6001	0.5845
		(0.0185)	(0.0180)	(0.0024)	(0.0119)	(0.0135)	(0.0062)
	0.60	0.5328	0.5449	0.6006	0.6239	0.6114	0.5832
		(0.0177)	(0.0162)	(0.0015)	(0.0149)	(0.0652)	(0.0063)
	0.75	0.6701	0.6864	0.7496	0.6551	0.6289	0.5838
		(0.0171)	(0.0145)	(0.0007)	(0.0202)	(0.0335)	(0.0064)
100	0.15	0.1323	0.1334	0.1525	0.5978	0.5971	0.5929
		(0.0095)	(0.0096)	(0.0015)	(0.0048)	(0.0049)	(0.0027)
	0.30	0.2774	0.2803	0.3008	0.5996	0.5977	0.5928
		(0.0097)	(0.0097)	(0.0014)	(0.0058)	(0.0060)	(0.0029)
	0.45	0.4219	0.4263	0.4507	0.6031	0.5994	0.5916
		(0.0092)	(0.0089)	(0.0012)	(0.0071)	(0.0075)	(0.0030)
	0.60	0.5645	0.5711	0.5998	0.6116	0.6053	0.5909
		(0.0081)	(0.0075)	(0.0008)	(0.0083)	(0.0088)	(0.0031)
	0.75	0.7085	0.7165	0.7496	0.6314	0.6197	0.5930
		(0.0066)	(0.0059)	(0.0003)	(0.0112)	(0.0121)	(0.0030)
200	0.15	0.1417	0.1424	0.1514	0.5986	0.5982	0.5965
		(0.0052)	(0.0052)	(0.0007)	(0.0024)	(0.0024)	(0.0013)
	0.30	0.2889	0.2903	0.3017	0.5997	0.5988	0.5959
		(0.0051)	(0.0051)	(0.0007)	(0.0029)	(0.0029)	(0.0014)
	0.45	0.4340	0.4364	0.4503	0.6025	0.6007	0.5956
		(0.0046)	(0.0045)	(0.0006)	(0.0034)	(0.0035)	(0.0014)
	0.60	0.5826	0.5855	0.5997	0.6051	0.6022	0.5956
		(0.0038)	(0.0036)	(0.0004)	(0.004)	(0.0045)	(0.0014)
	0.75	0.7303	0.7343	0.7500	0.6138	0.6079	0.5964
		(0.0028)	(0.0026)	(0.0002)	(0.0065)	(0.0067)	(0.0014)
400	0.15	0.1460	0.1463	0.1505	0.5993	0.5991	0.5984
		(0.0025)	(0.0025)	(0.0003)	(0.0012)	(0.0012)	(0.0007)
	0.30	0.2952	0.2959	0.3006	0.5993	0.5989	0.5979
		(0.0025)	(0.0025)	(0.0004)	(0.0014)	(0.0014)	(0.0007)
	0.45	0.4430	0.4441	0.4501	0.6008	0.6000	0.5985
		(0.0024)	(0.0024)	(0.0003)	(0.0019)	(0.0019)	(0.0007)
	0.60	0.5910	0.5925	0.6000	0.6039	0.6025	0.5989
		(0.0019)	(0.0018)	(0.0002)	(0.0023)	(0.0023)	(0.0007)
	0.75	0.7396	0.7414	0.7500	0.6074	0.6045	0.5981
		(0.0014)	(0.0013)	(0.0001)	(0.0036)	(0.0036)	(0.0007)

Table 4.5: counts of vagrancy data descriptive statistics.

Minimum	Median	Mean	Variance	Skewness	Kurtosis	$\widehat{\rho}(1)$	Maximum
1	1	1.3402	0.4358	2.2851	8.5345	0.1959	4

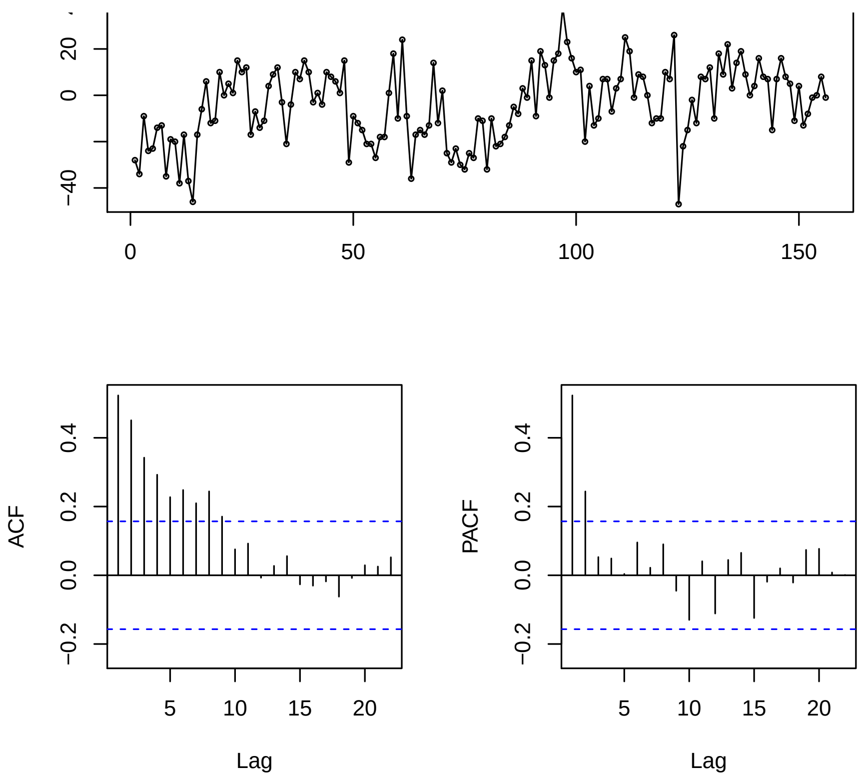


Figure 4.3: The time series counts of vagrancy, the autocorrelation function and the partial autocorrelation function.

Table 4.6: Estimated parameters (with corresponding standard errors in parentheses), RMS and AIC.

Model	CML estimates	RMS	AIC
BINAR(1)	$\widehat{\alpha} = 0.1156\ (0.0468)$ $\widehat{\lambda} = 0.1577\ (0.0493)$	0.6489	218.9168
Truncated Poisson INAR(1)	$\widehat{\alpha} = 0.0867\ (0.0600)$ $\widehat{\lambda} = 0.4225\ (0.1514)$	0.6507	221.2132
ZTPINAR(1)	$\widehat{\alpha} = 0.3886\ (0.1394)$ $\widehat{\lambda} = 0.6216\ (0.1025)$	0.6468	217.8503

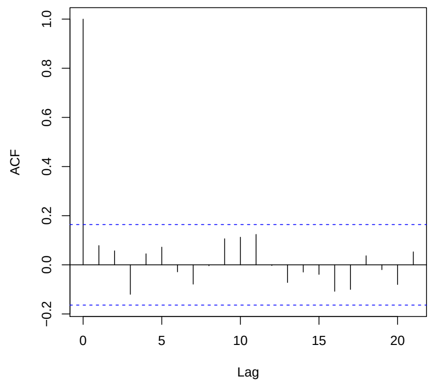


Figure 4.4: Sample autocorrelations of the residuals obtained from Borel INAR(1) model.

Furthermore, the behavior of the series indicates that it may be mean stationary. We compare our model with the shifted models: Truncated Poisson INAR(1) model (Bourguignon and Vasconcellos, 2015) and ZTPINAR(1) model (Bakouch and Ristić, 2010). Table 4.6 presents the CML estimates with corresponding standard errors in parentheses, Akaike information criterion (AIC) and the root mean square of differences between observed and predicted values (RMS). From Table 4.6, we observe that the BINAR(1), truncated Poisson INAR(1) and ZTPINAR(1) models are competitive. The BINAR(1) model showed better goodness-of-fit statistics compared with the truncated Poisson INAR(1) model. From the results, we conclude that the BINAR(1) is adequate for this time series.

The sample autocorrelations of the residuals from Borel INAR(1) process are displayed in Figure 4.4, confirming that a first order model is adequate to fit the data set.

Now, our interest is to test the null hypothesis $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda < \lambda_0$, where

Table 4.7: Results from hypothesis test applied to time series counts of vagrancy: $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda < \lambda_0$.

Data set size	$\widehat{\lambda}_{\text{CLS}}$	$\widehat{\alpha}_{\text{CLS}}$	s.d.$_{\widehat{\lambda}_{\text{CLS}}}$	test statistic	p-value	critical value
144	0.0733	0.1962	1.1469	−3.2290 1	0.0006	0.5082

$\lambda_0 = (3 - \sqrt{5})/2 = 0.382$ at significance level $\delta = 0.05$. In other words, we want to test the null hypothesis of an equidispersed model against an underdispersed model. For this, we considered the CLS estimator and the asymptotic distribution given in Theorem 4.4.1. In Table 4.7 we have the empirical results of the CLS estimators to λ and α, the standard deviation of the CLS estimator of λ, the test statistic, the p-value and the critical value.

With the results given in Table 4.7, we have that $\widehat{\lambda}_{\text{CLS}} = 0.0733 < 0.5082 = \lambda_0 + q_\delta\sqrt{\widehat{\sigma}^2_{\lambda_0}/144}$, therefore we reject the null hypothesis $H_0 : \lambda = \lambda_0$ in favour of the alternative hypothesis $H_1 : \lambda < \lambda_0$ at significance level $\delta = 0.05$. So, we have evidence for an underdispersed model, as expected. Alternatively, from a p-value analysis we have that p-value = 0.0006 $<$ 0.05 = δ, confirming the result of the hypothesis test.

4.6.2 *Modelling overdispersion*

The data set of the our second application consist of 144 observations are from Macdonald and Zucchini (1970). Consider the series of weekly sales (in integer units) of a particular soap product in a supermarket. The data are taken from a database provided by the Kilts Center for Marketing, Graduate School of Business of the University of Chicago,
:http://gbswww.uchicago.edu/kilts/research/db/dominicks (The product is 'Zest White Water 15 oz.', with code 3700031165). In order to use the BINAR model, as made by Bakouch and Ristić (2010), we transformed the series adding 1 to each observation.

Table 4.8: Series of weekly sales descriptive statistics.

Minimum	Median	Mean	Variance	Skewness	Kurtosis	$\widehat{\rho}(1)$	Maximum
1	5	6.4421	15.4012	1.4998	5.7926	0.3923	23

In Table 4.8 we presents some descriptive measures for the data set of weekly sales, which include central tendency statistics, variance, skewness and kurtosis, among others. We can observe that the mean is smaller than the variance, showing overdispersion.

In Figure 4.5, the time series data and their sample autocorrelation and partial autocorrelation are displayed. Figure 4.5 shows that a first order autoregressive model may be appropriate for the given data series. Furthermore, the behavior of the series indicates that it may be mean stationary. The sample autocorrelations of the residuals from Borel INAR(1) process are displayed in Figure 4.6.

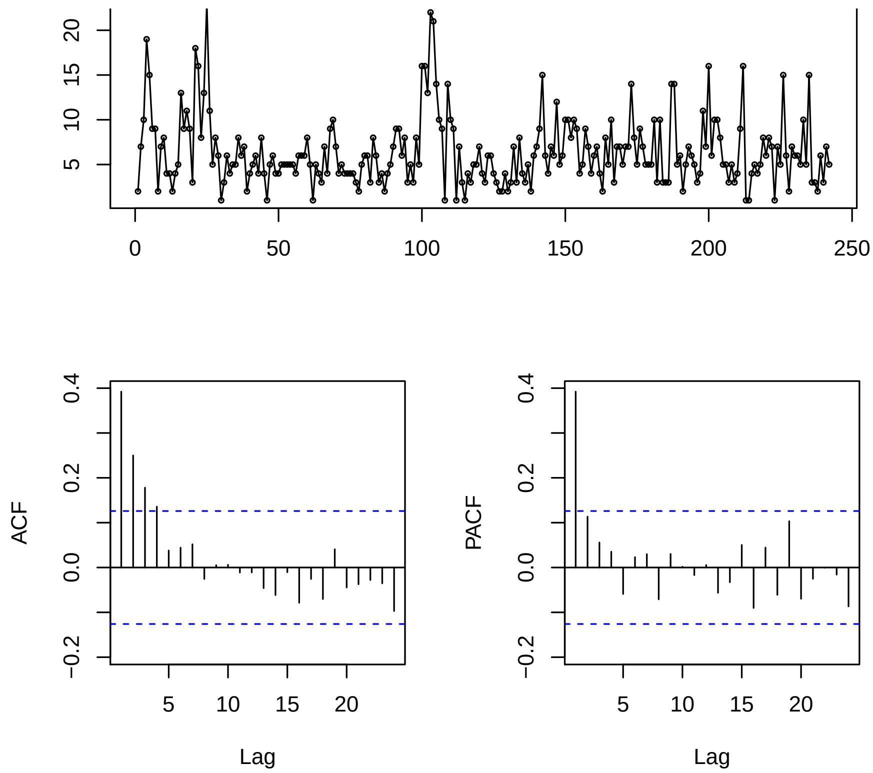


Figure 4.5: Time series of weekly sales, the autocorrelation function and the partial autocorrelation function.

We compare our model to the Truncated Poisson INAR(1) model (Bourguignon and Vasconcellos, 2015) and also to the ZTPINAR(1) model (Bakouch and Ristić, 2010). Table 4.9 presents the CML estimates with corresponding standard errors in parentheses, AIC and the RMS. From Table 4.9, we observe that the BINAR(1) model showed better goodness-of-fit statistics compared with the other models. From the results, we conclude that the BINAR(1) is adequate for this time series.

Our interest now lies in testing the null hypothesis $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda > \lambda_0$, where $\lambda_0 = (3 - \sqrt{5})/2 = 0.382$ on significance level $\delta = 0.05$, that is, we want test the null hypothesis of an equidispersed model against an overdispersed model. Again, we considered the CLS estimator to λ and the asymptotic distribution given in Theorem 4.4.1. In Table 4.10 we have the empirical results of the CLS estimator to λ and α, the standard deviation of CLS estimator, the test statistic, p-value and the critical value.

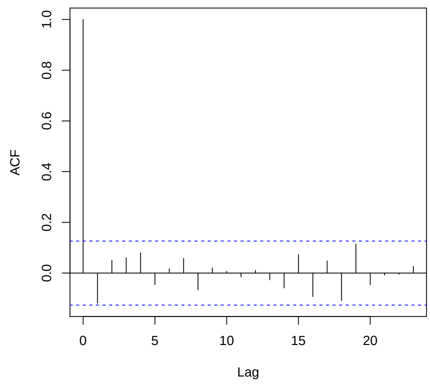


Figure 4.6: Sample autocorrelation of the residuals obtained from Borel INAR(1) model.

Table 4.9: Estimated parameters (with corresponding standard errors in parentheses), RMS and AIC.

Model	CML estimates	RMS	AIC
BINAR(1)	$\widehat{\alpha} = 0.4647\ (0.0211)$	3.6093	1317.036
	$\widehat{\lambda} = 0.7112\ (0.0305)$		
Truncated Poisson INAR(1)	$\widehat{\alpha} = 0.2222\ (0.0335)$	3.6597	1328.947
	$\widehat{\lambda} = 4.9934\ (0.2564)$		
ZTPINAR(1)	$\widehat{\alpha} = 0.2433\ (0.0361)$	3.6457	1324.429
	$\widehat{\lambda} = 6.4567\ (0.2109)$		

With the results given in Table 4.10, we have that $\widehat{\lambda}_{\mathrm{CLS}} = 0.7455 > 0.5118 = \lambda_0 + q_\delta\sqrt{\widehat{\sigma}^2_{\lambda_0}/144}$, therefore we reject the null hypothesis $H_0 : \lambda = \lambda_0$ in favour of the alternative hypothesis $H_1 : \lambda < \lambda_0$ at significance level $\delta = 0.05$. So, we have evidence for an overdispersed model, as expected. Alternatively, from a p-value analysis, we have a extremely small p-value = 4.5519 $\times 10^{-15}$ $<$ 0.05 = δ, confirming the result of the hypothesis test.

4.7 Concluding remarks

In this chapter we introduced the shifted first order integer-valued autoregressive process with Borel innovations. This model was based on the binomial thinning with Borel distribution innovations. The Borel INAR(1) model is suited to equidispersed, underdispersed and overdispersed data counts set. The basic properties of this model are obtained. The closed ex-

Table 4.10: Results from hypothesis test applied to time series weekly sales: $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda > \lambda_0$.

Data set size	$\widehat{\lambda}_{\text{CLS}}$	$\widehat{\alpha}_{\text{CLS}}$	s.d.$_{\widehat{\lambda}_{\text{CLS}}}$	test statistic	p-value	critical value
242	0.7455	0.3925	0.7295	7.7513	4.5519 $\times 10^{-15}$	0.5118

pressions from CLS and YW estimators of unknown parameters are derived and we obtain of their asymptotic distribution. An hypothesis test based on the asymptotic distribution of CLS estimators was formulated to test equidispersed against overdispersed or underdispersed model. We made a test power analysis by simulating the rejected rate for different sample sizes considering the true and estimated value of α and the asymptotic distribution. Two applications were presented showing that the Borel INAR(1) model is suited to shifted underdispersed and overdispersed data sets and the results of a hypothesis test confirmed the assessment.

4.8 Appendix

Proof of Theorem 4.4.1.

The Borel INAR(1) process $\{X_t\}_{t\in\mathbb{Z}}$ is a branching process with immigration whose offspring distribution is Bernoulli and the immigration process has Borel distribution, both having all moments finite. By Klimko and Nelson (1978), the CLS estimators of α and μ_ϵ are strongly consistent and their asymptotic joint distribution can be obtained applying the results from Section 5 in the same paper. We have,

$$\sqrt{T}(\widehat{\alpha}_{\text{CLS}} - \alpha, \widehat{\mu}_{\epsilon;\text{CLS}} - \mu_\epsilon) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}^*)$$

where $\widehat{\alpha}_{\text{CLS}}$ is give in (4.8) and $\widehat{\mu}_{\epsilon;\text{CLS}}$ is given by the expression

$$\widehat{\mu}_{\epsilon;\text{CLS}} = \frac{1}{T}\left(\sum_{t=1}^{T} X_t - \widehat{\alpha}_{\text{CLS}} \sum_{t=2}^{T} X_{t-1}\right),$$

with $\mathbf{\Sigma}^* = (\sigma^*_{ij})$ being obtained from the results given in Klimko and Nelson (1978), Section 5, by

$$
\begin{aligned}
\sigma_{11}^{*} &= (1-\alpha^2) + \frac{\alpha(1-\alpha)(1-\alpha^2)^2(1-\lambda)^5}{(1-\alpha^3)[\lambda+\alpha(1-\lambda)^2]^2}\left[\frac{\alpha}{1-\alpha} + \frac{3\alpha^3}{1+\alpha} + \frac{3\alpha^2\lambda}{(1+\alpha)(1-\lambda)^2} + \frac{1+2\lambda}{(1-\lambda)^4}\right],\\
\sigma_{12}^{*} &= \alpha(1-\alpha) - \frac{\lambda(1-\alpha)}{(1-\lambda)[\lambda+\alpha(1-\lambda)^2]} - \frac{\alpha(1+\alpha)(1-\lambda)}{\lambda+\alpha(1-\lambda)^2} - [(1+\alpha)\\
&\quad + 3\alpha^2(1-\alpha)]\frac{\alpha^2(1-\alpha^2)(1-\lambda)^5}{[\lambda+\alpha(1-\lambda)^2]^2} - [3\alpha^2\lambda(1-\lambda)^2 + (1+2\lambda)(1+\alpha)]\frac{\alpha(1-\alpha)^2(1+\alpha)(1-\lambda)}{[\lambda+\alpha(1-\lambda)^2]^2},\\
\sigma_{22}^{*} &= \frac{1}{(1-\lambda)^2(1-\alpha)^2} - \frac{\alpha}{1-\lambda} + \frac{\lambda}{(1-\lambda)^3} + \frac{\alpha(1+\alpha)}{(1-\alpha)[\lambda+\alpha(1-\lambda)^2]}\\
&\quad + \frac{\alpha(1-\alpha)(1+\alpha)^2(1-\lambda)^3}{(1-\alpha^3)[\lambda+\alpha(1-\lambda)^2]^2}\left[\frac{\alpha}{1-\alpha} + \frac{3\alpha^3}{1+\alpha} + \frac{3\lambda\alpha^2}{(1+\alpha)(1-\lambda)^2} + \frac{1+2\lambda}{(1-\lambda)^4}\right].
\end{aligned}
$$

Now, we define the function $f := \mathbb{R}^2_* \to \mathbb{R}^2$, as follows:

$$f = (f_1, f_2)^\top = (f_1(x_1, x_2), f_2(x_1, x_2)) := \big(x_1, 1 - \frac{1}{x_2}\big),$$

where $\mathbb{R}^2_* := \{(x_1, x_2) \in \mathbb{R}^2; x_2 \neq 0\}$. We have that the f function is continuous and differentiable at its domain. According to the delta method, considering $\boldsymbol{\theta} = (\alpha, \mu_\epsilon)$ and $\mathbf{Y_T} = (\widehat{\alpha}_{\mathrm{CLS}}, \widehat{\mu}_{\epsilon;\mathrm{CLS}})$, we have

$$\sqrt{T}(f(\mathbf{Y_T}) - f(\theta)) \xrightarrow{d} N(\mathbf{0}, \mathbf{D}\boldsymbol{\Sigma}^*\mathbf{D}^\top),$$

where $\mathbf{D}$ is the Jacobian matrix of f with respect to $\boldsymbol{\theta}$.

The first-order partial derivatives of f is given by

$$\frac{\partial f_1(x_1,x_2)}{\partial x_1} = 1, \quad \frac{\partial f_1(x_1,x_2)}{\partial x_2} = 0, \quad \frac{\partial f_2(x_1,x_2)}{\partial x_1} = 0, \quad \frac{\partial f_2(x_1,x_2)}{\partial x_2} = \frac{1}{x_2^2},$$

thus, the Jacobian matrix of f with respect to $\boldsymbol{\theta}$ is

$$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\mu_\epsilon^2} \end{pmatrix}.$$

We have

$$\boldsymbol{\Sigma} = \mathbf{D}\boldsymbol{\Sigma}^*\mathbf{D}^\top = \begin{pmatrix} \sigma_{11}^* & \frac{1}{\mu_\epsilon^2}\sigma_{12}^* \\ \frac{1}{\mu_\epsilon^2}\sigma_{12}^* & \frac{1}{\mu_\epsilon^4\sigma_{22}^*} \end{pmatrix}.$$

Finally, we obtain

$$\begin{aligned}
\sigma_{11} &= (1-\alpha^2) + \frac{\alpha(1-\alpha)(1-\alpha^2)^2(1-\lambda)^5}{(1-\alpha^3)[\lambda+\alpha(1-\lambda)^2]^2}\left[\frac{\alpha}{1-\alpha} + \frac{3\alpha^3}{1+\alpha} + \frac{3\alpha^2\lambda}{(1+\alpha)(1-\lambda)^2} + \frac{1+2\lambda}{(1-\lambda)^4}\right],\\
\sigma_{12} &= \alpha(1-\alpha)(1-\lambda)^2 - \frac{\lambda(1-\alpha)(1-\lambda)}{\lambda+\alpha(1-\lambda)^2} - \frac{\alpha(1+\alpha)(1-\lambda)^3}{\lambda+\alpha(1-\lambda)^2} - [(1+\alpha)\\
&\quad + 3\alpha^2(1-\alpha)]\frac{\alpha^2(1-\alpha^2)(1-\lambda)^7}{[\lambda+\alpha(1-\lambda)^2]^2} - [3\alpha^2\lambda(1-\lambda)^2 + (1+2\lambda)(1+\alpha)]\frac{\alpha(1-\alpha)^2(1+\alpha)(1-\lambda)^3}{[\lambda+\alpha(1-\lambda)^2]^2},\\
\sigma_{22} &= \{1-\alpha[\alpha\lambda + (1-2\alpha)(1+\lambda^2) + \alpha^2(1-\lambda)^2]\}\frac{1-\lambda}{(1-\alpha)^2} + \frac{\alpha(1+\alpha)(1-\lambda)^4}{(1-\alpha)(\lambda+\alpha(1-\lambda)^2)}\\
&\quad + \frac{3\alpha^3(1-\lambda)^5(1-\alpha^2)}{(1-\alpha^3)(\lambda+\alpha(1-\lambda)^2)} + [(1-\alpha)(1+2\lambda) + \alpha(1-\lambda)^4]\frac{\alpha(1+\alpha)^2(1-\lambda)^3}{(1-\alpha^3)[\lambda+\alpha(1-\lambda)^2]^2}.
\end{aligned}$$

Proof of Theorem 4.4.2.

The proof is done in the same way of Freeland and McCabe (2005). The proof to show that $\sqrt{T}(\widehat{\alpha}_{\text{CLS}} - \widehat{\alpha}_{\text{YW}})$ is $o_p(1)$ in Borel INAR(1) process is identically the proof in Poisson INAR(1) process.

Let

$$D_{\text{CLS}} = \frac{1}{T}\left[\sum_{t=2}^{T} X_{t-1}^2 - \frac{1}{T-1}\Big(\sum_{t=2}^{T} X_{t-1}\Big)^2\right] \text{ and } D_{\text{YW}} = \frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})^2 = \frac{1}{T}\sum_{t=1}^{T} X_t^2 - \bar{X}^2.$$

Thus,

$$\begin{aligned}
&\sqrt{T}(\widehat{\alpha}_{\text{YW}} - \widehat{\alpha}_{\text{CLS}}) =\\
&\sqrt{T}\left[\frac{D_{CLS}(\frac{1}{T}\sum_{t=2}^{T} X_t X_{t-1} - \bar{X}\frac{1}{T}\sum_{t=2}^{T} X_t - \bar{X}\frac{1}{T}\sum_{t=2}^{T} X_{t-1} + (1-\frac{1}{T})\bar{X}^2)}{D_{\text{CLS}}D_{\text{YW}}}\right]\\
&-\sqrt{T}\left[\frac{D_{\text{YW}}(\frac{1}{T}\sum_{t=2}^{T} X_t X_{t-1} - \frac{1}{T(T-1)}\sum_{t=2}^{T} X_t \sum_{t=2}^{T} X_{t-1})}{D_{\text{CLS}}D_{\text{YW}}}\right],
\end{aligned}$$

After some algebra, we conclude

$$\begin{aligned}
&\sqrt{T}(\widehat{\alpha}_{\text{YW}} - \widehat{\alpha}_{\text{CLS}}) =\\
&\frac{(D_{\text{CLS}} - D_{\text{YW}})}{D_{\text{CLS}}D_{\text{YW}}} \times \frac{\sum_{t=2}^{T} X_t X_{t-1}}{\sqrt{T}} - \frac{(D_{\text{CLS}}\sum_{t=2}^{T}\frac{X_t}{T} - D_{\text{YW}}\sum_{t=2}^{T}\frac{X_t}{T-1})}{D_{\text{CLS}}D_{\text{YW}}} \cdot \frac{\sum_{t=1}^{T} X_t}{\sqrt{T}}\\
&-\frac{D_{\text{YW}}}{D_{\text{CLS}}D_{\text{YW}}} \cdot \sum_{t=2}^{T}\frac{X_t}{T-1}\frac{X_T}{\sqrt{T}} + \frac{D_{\text{CLS}}}{D_{\text{CLS}}D_{\text{YW}}} \times \bar{X}\frac{(X_T - \bar{X})}{\sqrt{T}}\\
&= o_p(1)O_p(1) - o_p(1)O_p(1) - O_p(1)O_p(1)o_p(1) + O_p(1)O_p(1)o_p(1) = o_p(1).
\end{aligned}$$

For λ estimators, define

$$D_{\text{CLS}} = \frac{1}{T}\sum_{t=2}^{T} X_t - \widehat{\alpha}_{\text{CLS}}\frac{1}{T}\sum_{t=2}^{T} X_{t-1} \quad \text{and} \quad D_{\text{YW}} = \overline{X}(1-\widehat{\alpha}_{\text{YW}}).$$

Note that D_{CLS} and D_{YW} converges to the same non zero constant $\mu(1-\alpha)$.

We have

$$\begin{aligned}
\sqrt{T}(\widehat{\lambda}_{\text{YW}} - \widehat{\lambda}_{\text{CLS}}) &= \sqrt{T}\left[\frac{1}{\overline{X}(1-\widehat{\alpha}_{\text{YW}})} - \frac{T-1}{\sum_{t=2}^{T} X_t - \widehat{\alpha}_{\text{CLS}}\sum_{t=2}^{T} X_{t-1}}\right] \\
&= \sqrt{T}\left\{\frac{\frac{1}{T}\sum_{t=2}^{T} X_t - \widehat{\alpha}_{\text{CLS}}\frac{1}{T}\sum_{t=2}^{T} X_{t-1} - (1-\frac{1}{T})[\overline{X}(1-\widehat{\alpha}_{\text{YW}})]}{D_{\text{CLS}}D_{\text{YW}}}\right\}.
\end{aligned}$$

After some algebra, we have

$$\begin{aligned}
\sqrt{T}(\widehat{\lambda}_{\text{YW}} - \widehat{\lambda}_{\text{CLS}}) &= \frac{\frac{1}{\sqrt{T}}(X_T\widehat{\alpha}_{\text{CLS}} - X_1) - \sqrt{T}(\widehat{\alpha}_{\text{CLS}} - \widehat{\alpha}_{\text{YW}})\overline{X} + \frac{\overline{X}}{\sqrt{T}}(1-\widehat{\alpha}_{\text{YW}})}{D_{\text{CLS}}D_{\text{YW}}} \\
&= o_p(1) - o_p(1) + o_p(1) = o_p(1).
\end{aligned}$$

References

[1] Al-Osh, M.A., Alzaid, A.A. (1987). First-order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, **8**, 261–275.

[2] Al-Osh, M. A. and Alzaid, A. A. (1991). Binomial autoregressive moving average models. *Stochastic Models*, **7**, 261-282.

[3] Al-Osh, M. A. and Aly, E. E. A. (1992). First order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics-Theory and Methods*, **21**, 2483-2492.

[4] Al-Nachawati, H., Alwasel, I., Alzaid, A. A. (1997). Estimating the parameters of the generalized Poisson AR(1) process. *Journal of Statistical Computation and Simulation*, **56**, 337-352.

[5] Alzaid, A. A. and Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR (p)) process. *Journal of Applied Probability*, **27**, 314-324.

[6] Alzaid, A. A. and Al-Osh, M. A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics*, **45**, 223-232.

[7] Alzaid, A. A., Omair, M. A. (2014). Poisson difference integer valued autoregressive model of order one. *Bulletin of the Malaysian Mathematical Sciences Society*, **37**, 465-485.

[8] Awale, M., Ramanathan, T. V. and Kale, M. (2017). Coherent forecasting in integer-valued AR(1) models with geometric marginals. *Journal of Data Science*, **15**, 95-114.

[9] Bakouch, H. S. (2010). Higher-order moments, cumulants and spectral densities of the NGINAR (1) process. *Statistical Methodology*, **7**, 1-21.

[10] Bakouch, H. S., Ristić, M. M. (2010). Zero truncated Poisson integer-valued AR(1) model. *Metrika*, **72**, 265–280.

[11] Barreto-Souza, W., Bourguignon, M. (2013). A skew true INAR(1) process with application. *AStA - Advances in Statistical Analysis*, **99**, 189-208.

[12] Billingsley, P. Probability and Measure. John Wiley & Sons, New York, 1979.

[13] Borel, E. (1942). Sur l'emploi du théoreme de Bernoulli pour faciliter le calcul d'une infinité de coefficients. Application au probleme de l'attentea un guichet. *CR Acad. Sci. Paris*, **214**, 452–456.

[14] Bourguignon, M. and Vasconcellos, K. L. P. (2015a). Improved estimation for Poisson INAR(1) models. *Journal of Statistical Computation and Simulation*, **85**, 2425-2441.

[15] Bourguignon, M., Vasconcellos, K. L. (2015b). First order non-negative integer valued autoregressive processes with power series innovations. *Brazilian Journal of Probability and Statistics*, **29**, 71–93.

[16] Bourguignon, M., Vasconcellos, K. L. (2016). A new skew integer valued time series process. *Statistical Methodology*, **31**, 8-19.

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

[18] Box, G. E., Jenkins, G. M. and Reinsel, G. C. (1994). Time Series Analysis: Forecasting and Control. Prentice-Hall, New Jersey.

[19] Chesneau, C., Kachour, M. (2012). A parametric study for the first-order signed integer-valued autoregressive process. *Journal of Statistical Theory and Practice*, **6**, 760-782.

[20] Consul, P.C. (1974). A simple urn model dependent upon predetermined strategy. Sankhy$\bar{a}$: *The Indian journal of Statistics*. series B, **36**, 391-399.

[21] Consul, P. C. (1975). On a characterization of Lagrangian Poisson and quasi-binomial distributions. *Communications in Statistics-Theory and Methods*, **4**, 555-563.

[22] Consul, P.C. (1986). On the differences of two generalized Poisson variates. *Communications in Statistics - Simulation and Computation*, **15**, 761-767.

[23] Consul, P. C. (1989). Generalized Poisson Distribution: Properties and Applications. Marcel Dekker Inc., New York.

[24] Consul, P.C. (1990). On some properties and applications of quasi-binomial distribution. *Communications in Statistics - Theory and Methods*, **19**, 477-504.

[25] Consul, P.C., Famoye, F. (2006). Lagrangian Probability Distributions. Birkhäuser Boston.

[26] Consul, P. C., Jain, G. C. (1973). A generalization of Poisson distribution, *Technometrics*, **15**, 791-799.

[27] Consul, P. C., Mittal, S. P. (1975). A new urn model with predetermined strategy, *Biometrical Journal*, **17**, 67-75.

[28] Du, J., Li, Y. (1991). The integer valued autoregressive (INAR(p)) model, *Journal of Times Series Analysis*, **12**, 129–142.

[29] Fazal, S.S. (1976). A test for quasi-binomial distribution. *Biometrical Journal*, **18**, 619-622.

[30] Freeland, R. K., McCabe, B. (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics and Probability Letters*, **73**, 147–153.

[31] Freeland, R.K. (2010). True integer value time series. *AStA Advances in Statistical Analysis*, **94**, 217-229.

[32] Haight, F. A., Breuer, M. A. (1960). The Borel-Tanner distribution. *Biometrika*, **47**, 143–150.

[33] Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. CORE discussion paper, 2003-63, University of Louvain, Belgium.

[34] Joe, H., Zhu, R. (2005). Generalized Poisson Distribution: the property of mixture of Poisson and comparison with Negative-binomial distribution, *Biometrical Journal*, **47**, 219-229.

[35] Johnson,N.L., Kotz, S., Kemp, A. W. (2005). Univariate Discrete Distributions. John Wiley & Sons, Inc., New York. third edition.

[36] Jung, R. C., Ronning, G. and Tremayne, A. R. (2005). Estimation in conditional first order autoregression with discrete support. *Statistical Papers*, **46**, 195-224.

[37] Jung, R. C., Tremayne, A. R. (2006). Binomial thinning models for integer time series. *Statistical Modelling*, **6**, 81–96.

[38] Karlis, D., Ntzoufras, I. (2006). Bayesian analysis of the differences of count data. *Statistics in Medicine*, **25**, 1885-1905.

[39] Klimko, L. A., Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 629–642.

[40] Korwar, R.M. (1977). On characterizing lagrangian poisson and quasi-binomial distributions. *communications in Statistics - Theory and Methods*, **6**, 1409-1415.

[41] Nastić, A. S., Ristić, M. M., Djordjević, M. S. (2016). An INAR model with discrete Laplace marginal distributions. *Brazilian Journal of Probability and Statistics*, **30**, 107-126.

[42] MacDonald, I. L., and Zucchini, W. (1997). Hidden Markov and other models for discrete-valued time series. CRC Press.

[43] McKenzie, E. (1985). Some simple models for discrete variate time series. *JAWRA Journal of the American Water Resources Association*, **21**, 645–650.

[44] McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, **18**, 679-705.

[45] McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, **20**, 822-835.

[46] Nastić, A. S. (2012). On shifted geometric INAR(1) models based on geometric counting series. *Communications in Statistics-Theory and Methods*, **41**, 4285–4301.

[47] 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.

[48] Shahtahmassebi, G., Moyeed, R. (2014). Bayesian modelling of integer data using the generalized Poisson difference distribution. *Int. Jornal of Statistics and Probability,* **3**, 35-48.

[49] Shahtahmassebi, G., Moyeed, R. (2016). An application of the generalized Poisson difference distribution to the Bayesian modelling of football scores. *Statistica Neerlandica*, **70**, 260-273.

[50] Schweer, S., Weiß, C. H. (2014). Compound Poisson INAR(1) processes: stochastic properties and testing for overdispersion. *Computational Statistics and Data Analysis*, **77**, 267 – 284.

[51] Shenton, L. R. (2006). Quasibinomial distributions. *Encyclopedia of Statistical Sciences*, **10**, 6744-6746.

[52] Silva, M.E., Oliveira, V. L. (2004). Difference equations for the higher-order moments and cumulants of the INAR(1) model. *Journal of Time Series Analysis*, **25**, 317–333.

[53] Steutel, F. W.; van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 893-899.

[54] Silva, I. M. M. D. (2005). Contributions to the Analysis of Discrete-valued Time Series. PhD Thesis, University of Porto, Portugal.

[55] Steutel, F. W.; van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 893–899.

[56] Weiß, C. H. (2007). Controlling correlated processes of Poisson counts. *Quality and Reliability Engineering International*, **23**, 741–754.

[57] Weiß , C. H. (2008). Thinning operations for modelling time series of counts-a survey. *AStA Advances in Statistical Analysis*, **92**, 319-341.

[58] Weiß , C. H. (2009). Modelling time series of counts with overdispersion. *Statistical Methods and Applications*, **18**, 507-519.

[59] Weiß, C. H. and Kim, H. Y. (2011). Binomial AR (1) processes: moments, cumulants, and estimation. *Statistics*, **47**, 494-510.

[60] Weiß, C. H. (2013). Integer-valued autoregressive models for counts showing underdispersion. *Journal of Applied Statistics*, **40**, 1931–1948.

[61] Weiß, C. H. and Kim, H. Y. (2013). Parameter estimation for binomial AR (1) models with applications in finance and industry. *Statistical Papers*, **54**, 563-590.

[62] Weiß, C. H. and Schweer, S. (2016). Bias corrections for moment estimators in Poisson INAR (1) and INARCH (1) processes. *Statistics and Probability Letters*, **112**, 124-130.