FEDERAL UNIVERSITY OF PERNAMBUCO
CENTRE FOR NATURAL AND EXACT SCIENCES
GRADUATE PROGRAM IN STATISTICS

MODELLING AND INFERENCE IN INTEGER-VALUED TIME SERIES

MARCELO BOURGUIGNON PEREIRA

Doctoral thesis

Recife
2014

Federal University of Pernambuco
Centre for Natural and Exact Sciences

Marcelo Bourguignon Pereira

MODELLING AND INFERENCE IN INTEGER-VALUED 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 Ph.D. in Statistics.

Advisor: Professor Dr. Klaus Leite Pinto Vasconcellos

Recife
2014

Catalogação na fonte
Bibliotecária Jane Souto Maior, CRB4-571

P436m Pereira, Marcelo Bourguignon
Modelling and inference in integer-valued time series. / Marcelo Bourguignon Pereira. – Recife: O Autor, 2014.
97 f.: il., fig., tab.

Orientador: Klaus Leite Pinto Vasconcellos.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, 2014.
Inclui referências e apêndices.

1. Séries temporais. I. Vasconcellos, Klaus Leite Pinto (orientador). II. Título.

519.232 CDD (23. ed.) UFPE- MEI 2014-187

MARCELO BOURGUIGNON PEREIRA

MODELLING AND INFERENCE IN INTEGER-VALUED 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: 15/12/2014.

BANCA EXAMINADORA

____
Prof. PhD. Klaus Leite Pinto Vasconcellos (Orientador)
Universidade Federal de Pernambuco

____
Prof. PhD. Francisco Cribari Neto (Examinador Interno)
Universidade Federal de Pernambuco

____
Prof. PhD. Gauss Moutinho Cordeiro (Examinador Interno)
Universidade Federal de Pernambuco

____
Prof.ª Doutora Luz Milena Zea Fernández (Examinadora Externa)
Universidade Federal do Rio Grande do Norte

____
Prof. PhD. Renato Martins Assunção (Examinador Externo)
Universidade Federal de Minas Gerais

Dedico este trabalho a minha mãe Orly e meu pai Alberto.

Agradecimentos

Quero agradecer em primeiro lugar a Deus, fonte de força inesgotável por ter me dado força, perseverança e paciência.

Agradeço muito especialmente aos meus pais pelo interesse em me ver concluir o curso e por serem minha eterna inspiração para continuar na luta por um objetivo na vida.

Ao meu orientador Professor Dr. Klaus Leite Pinto Vasconcellos pela orientação, paciência, ensinamentos, dedicação, sugestões e valiosas recomendações que tornaram possível este trabalho.

Ao professor Gauss Moutinho Cordeiro pelos ensinamentos, pelas oportunidades, pelas pesquisas e por todo apoio. O professor Gauss não foi meu orientador no doutorado, mas também o considero como orientador e amigo.

Aos meus amigos de Pernambuco, com destaque para os amigos Rodrigo, Agrinaldo, Luz, Josivandro, Dona Dada, Waldemar, Manoel, Cícero, Pedro, Poema, Fernanda e Laércio, que se tornaram a minha família nesses anos de mestrado/doutorado e com certeza foram fundamentais para a conclusão do doutorado, gostaria de expressar minha profunda gratidão.

A Universidade Federal do Piauí, em especial, aos professores do departamento de estatística: Valmária, Fernando, Cleide, Lya, Ronei, Rita, Dani, Max, Jeremias, Atila, Keliny, Jackelya e Eldo.

A Valéria Bittencourt, secretária da pós-graduação em estatística, pela competência, carinho, amizade e atenção. Obrigado por sempre me ajudar em tudo.

A minha esposa Gleici pela paciência, apoio e incentivo constantes.

À CAPES, pelo apoio financeiro.

Não sou nada

Nunca serei nada

Não posso querer ser nada

À parte isso, tenho em mim todos os sonhos do mundo.

(Fernando Pessoa).

Resumo

Séries temporais de valores inteiros têm chamado a atenção pela importância em aplicações nas diversas áreas de conhecimento. Os processos estocásticos usuais assumem que as marginais são contínuas e, em geral, não são adequados para modelar séries de contagem. Portanto, surge a necessidade de investigar metodologias apropriadas para séries temporais com distribuições marginais discretas. Em particular, o estudo de novos modelos e o estudo de estimadores corrigidos no contexto de processos de valores inteiros motivam uma vertente de pesquisa de grande interesse para aplicações práticas e são os principais objetivos desta pesquisa. Este trabalho está dividido em quatro capítulos independentes, todos no contexto de séries temporais de valores inteiros. No segundo capítulo, introduzimos o processo autorregressivo de primeira ordem para valores inteiros não negativos com inovações de séries de potência. As principais propriedades do modelo proposto são derivadas, tais como os momentos, a função de autocovariância e a função de autocorrelação. Discutimos a estimação dos parâmetros do processo pelos métodos de Yule-Walker, mínimos quadrados condicionais e máxima verossimilhança condicional. Ensaios de Monte Carlo são realizados para comparar os vícios e erros quadráticos médios dos três estimadores para os parâmetros do modelo proposto. Ilustramos a utilidade do novo processo através de duas aplicações a conjuntos de dados reais. No terceiro capítulo, consideramos o processo Poisson INAR(1) e obtivemos o viés de segunda ordem do estimador diferença de quadrados e utilizamos esse viés para definir um estimador com viés reduzido. O comportamento do estimador de mínimos quadrados condicional corrigido também é estudado. Além disso, encontramos a distribuição assintótica dos estimadores propostos e comparamos os desempenhos dos estimadores propostos por meio de um extensivo estudo de simulação. Um uso prático deste modelo e das fórmulas obtidas para correção de viés é apresentado. No capítulo 4, propomos dois novos estimadores para o processo Poisson INAR(1), os quais são robustos na presença de observações atípicas ou outliers. Resultados de Monte Carlo mostraram que, em geral, os estimadores propostos para os parâmetros do processo Poisson INAR(1) são robustos na presença de outliers aditivos. Os estimadores propostos são aplicados a um conjunto de dados real. No quinto capítulo, introduzimos o processo autorregressivo de primeira ordem para valores inteiros com distribuição marginal geométrica-Poisson. O novo processo permite valores negativos para a série. Várias propriedades do processo são estabelecidas. Os parâmetros desconhecidos do modelo são estimados utilizando o método de Yule-Walker e suas propriedades assintóticas são consideradas. Alguns resultados numéricos dos estimadores são apresentados com uma discussão sobre os resultados obtidos. A flexibilidade do novo modelo é ilustrada com uma aplicação a um conjunto de dados reais.

Palavras-chave: Correção de viés. Distribuição séries de potência. Estimador de máxima verossimilhança condicional. Estimador de mínimos quadrados condicional. Estimador de Yule-Walker. Operador thinning. Outliers aditivos; Processos INAR(1).

Abstract

Time series of counts have been the focus of attention due to their importance in several areas of knowledge. The usual stochastic processes assume continuous marginal distributions and therefore are not suitable for modeling series of counts. Thus arises the need to investigate methodologies for time series with discrete marginal distributions. In particular, the study of new models and the behavior of corrected estimators of the integer-valued processes motivates an important research area with practical applications and is the main objective of this work. This work is divided into four independent chapters, all in the context of integer-valued time series. In the second chapter, we introduce a first order non-negative integer-valued autoregressive process with power series innovations. The main properties of the model are derived such as the mean, variance and autocorrelation function. The Yule-Walker, conditional least squares and conditional maximum likelihood estimators of the model parameters are obtained. An extensive Monte Carlo experiment is conducted to evaluate the performance of these estimators in finite samples. Applications to two real data sets are given to show the flexibility and potentiality of the new model. In the second third, we consider the first-order Poisson autoregressive model, which may be suitable in situations where the time series data are non-negative integer-valued. We derive the second order bias of the squared difference estimator for one of the parameters and use it to define a bias-adjusted estimator. The behavior of a modified conditional least squares estimator is also studied. Further, we assess the asymptotic properties of the estimators here discussed. We present numerical evidence, based on Monte Carlo simulation studies, that the proposed bias-adjusted estimator outperforms all other estimators in small samples. We also present an application to a real data set. In the fourth chapter, methods based on ranks for estimating the parameters of the Poisson INAR(1) model in the presence of additive outliers are proposed. The effects of additive outliers on the parameters estimates of the integer-valued time series are examined. Some numerical results of the estimators are presented with a discussion of the obtained results. We illustrate the usefulness of the proposed methods by means of an application to a real data set. The results presented motivate the use of the proposed methodology in practical situations in which a Poisson INAR(1) process contains additive outliers. In the fifth chapter, we introduce a stationary first-order integer-valued autoregressive process with geometric-Poisson marginals. The new process allows negative values for the series. Several properties of the process are established. The unknown parameters of the model are estimated by using the Yule-Walker method and their asymptotic properties are considered. Some numerical results of the estimators are presented with a discussion of the obtained results. The flexibility of the new model is illustrated with an application to a real data set.

Keywords: Additive outliers. Bias corrected. Conditional least squares estimator. Conditional maximum likelihood estimator. INAR(1) process. Power series distribution. Robust estimation. Thinning operator. Yule-Walker estimator.

List of Figures

List of Tables

Contents

CHAPTER 1

Introduction

Information generated from natural events, economic indicators, consequences of the development of a society, among other facts, requires a sophisticated data analysis to understand the dynamics of such phenomena and allows researchers to make inferences that enable decision making with some precision. In several areas of knowledge there are phenomena of interest which develop and change over time. The analysis of the collected data at different times, in general, leads to new problems that do not have standard methodologies. A simple introduction of correlation by sampling adjacent points in time can severely restrict the applicability of more conventional statistical methods, which assume independence of observations. The time series analysis consists of a set of appropriate tools to work in the context of correlated observations in time.

In the literature, a time series is defined as a set of observations y_t, each being collected at a specific time t. In order to characterize the nature of these observations, it is usual to assume that each value of the time series y_t is a realization of a random variable Y_t. That is, a time series y_t is then a realization of a family of random variables or of a stochastic process Y_t. A time series is said to be continuous when the observations are made continuously in time, i.e., if the set $\mathbb{T}$ of times at which observations are made is continuous. On the other hand, a time series is discrete if the set $\mathbb{T}$ of times at which observations are made is discrete. In this work, we consider discrete time series with $\mathbb{T} = \mathbb{Z}$. Note that these terms do not refer to the observed variable y_t, which can assume discrete or continuous values. When the random variables that compose the process follow a continuous distribution, the process is said to have a continuous marginal as, for example, the autoregressive moving average (ARMA) processes (Box et al., 1994). If the random variables that compose the process follow a discrete distribution, then the process is said to have a discrete marginal and is known as counting or integer-valued process.

Developed methodologies for discrete time series with continuous marginals have been considered for modeling series of counts. Typical models of time series, such as the ARMA assume that the marginal distribution is continuous and usually normally distributed. In some cases, continuous marginal models may be a suitable approximation. However, this assumption may be inappropriate for modeling time series with discrete marginal distributions. Therefore, there is a need to investigate appropriate methodologies for time series recorded in discrete time with discrete marginal distributions.

Despite the growing recognition of the need to model and simulate counting processes, few studies were published about this type of process until the end of the seventies. Over the past three decades, there have been several studies to develop a class of models suitable for this type of process. Currently, the study of counting processes has been a major highlight area in different fields.

Integer-valued time series have been the focus of attention mainly in medicine and economics (Franke and Seligmann, 1993, Freeland and McCabe, 2004). In both areas, time series of counts appear naturally. In medicine, the number of people hospitalized for epidemiological reasons or the number of admissions for respiratory causes, among others, are typical examples of series of counts. In economics, the study of numbers of shares sold in a period of time, patents, ship accidents, air accidents, stolen cars, passengers carried by a company, among other examples, motivate researchers to seek an appropriate statistical treatment for their analysis.

Models for count data have been widely used in several areas of study for modeling various phenomena. A fairly broad class of models is that based on the idea of "thinning" operator. Such models are discussed by McKenzie (1985, 1986, 1987, 1988), Al-Osh and Alzaid (1987, 1988, 1991) , Alzaid and Al-Osh (1988, 1990, 1993) , Du and Li (1991) and Al-Osh and Aly (1992). In general, detailed studies have been conducted not only on the formulation of models but also on properties (Silva and Oliveira, 2004), estimation (Jung, Ronning and Tremayne, 2005), statistical tests (Jung and Tremayne, 2003) and asymptotic distributions of model estimators (Freeland and McCabe, 2005) for different discrete marginal distributions.

This thesis is divided into four parts, composed by four 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 short overview of the four chapters is presented below.

The first order non-negative integer-valued autoregressive process with power series innovations based on the binomial thinning is introduced in Chapter 2. Various properties including second order moments, stationarity and the autocorrelation function are established for this new model. The Yule-Walker, conditional least squares and conditional maximum likelihood methods of estimation are used to estimate the model parameters. An extensive Monte Carlo experiment is conducted to evaluate the performances of these estimators in finite samples. Special sub-models are studied in some detail. Moreover, two real data sets are analyzed using the new model.

In Chapter 3, we study and reduce the biases of the parameters estimates for the Poisson

INAR(1) model. We derive the second order bias of the squared difference estimator for one of the parameters and use that bias to define a bias-adjusted estimator. The behavior of a modified conditional least squares estimator is also studied. Further, we access the asymptotic properties of the estimators here discussed. We present numerical evidence of the superiority of our estimator in terms of bias, based upon Monte Carlo simulation studies. We also present an application to a real data set.

In Chapter 4, methods based on ranks for estimating the parameters of the first-order integer-valued autoregressive model in the presence of additive outliers are proposed. In particular, we use the robust sample autocorrelations based on ranks to obtain estimators for the parameters of the Poisson INAR(1) process. The effects of additive outliers on the estimates of parameters of integer-valued time series are examined. Some numerical results of the estimators are presented with discussion. The proposed methods are applied to a dataset. The results presented here give motivation to use the methodology in practical situations in which a Poisson INAR(1) process contains additive outliers.

Finally, in Chapter 5, we introduce a stationary first-order integer-valued autoregressive process with geometric-Poisson marginals. The new process allows negative values for the series. Various properties including second order moments, stationarity and ergodicity are established for this new model. The unknown parameters of the model are estimated using the Yule-Walker method and their asymptotic properties are considered. Some numerical results for the estimators are presented with a discussion of the obtained results. The flexibility of the new model is illustrated with an application to a real data set.

References

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

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

Al-Osh, M.A. and Alzaid, A.A. (1988). Integer-valued moving average (INMA) process . *Statistical Papers*, **29**, 281–300.

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

Alzaid, A.A. and Al-Osh, M.A. (1988). First-order integer-valued autoregressive (INAR(1)) process.: distributional and regression properties. *Statistica Neerlandica*, **42**, 53–61.

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

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

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). *Time series analysis: forecasting and control*. Prentice-Hall, New Jersey.

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

Franke,J. and Seligmann, T. (1993). Conditional maximum-likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In *Developments in Time Series* (ed. T. Subba Rao). London: Chapman & Hall, 310–330.

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

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

Jung, R.C. and Tremayne, A.R. (2003). Testing for serial dependence in time series models of counts. *Journal of Time Series Analysis*, **24**, 65–84.

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

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, **21**, 645–650.

McKenzie, E. (1986). Autoregressive Moving-Average Processes with Negative- Binomial and Geometric Marginal Distributions. *Advances in Applied Probability*, **18**, 679–705.

McKenzie, E. (1987). Innovation distributions for gamma and negative binomial autoregressions. *Scandinavian Journal of Statistics*, **14**, 79–85.

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

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

CHAPTER 2

First order non-negative integer valued autoregressive processes with power series innovations

Resumo

Neste capítulo, introduzimos o processo autorregressivo de primeira ordem para valores inteiros não negativos com inovações séries de potência baseado no "thinning" binomial. Este novo modelo contém como casos particulares diversos modelos, como o modelo Poisson INAR-(1) (Al-Osh e Alzaid, 1987), o modelo geométrico INAR(1) (Jazi et al., 2012a) e muitos outros. As principais propriedades do modelo proposto são derivadas, tais como os momentos, a função de autocovariância e a função de autocorrelação. Discutimos a estimativa dos parâmetros do processo pelos métodos de Yule-Walker, mínimo quadrados condicionais e máxima verossimilhança condicional. Ensaios de Monte Carlo são realizados para comparar os vícios e erros quadráticos médios dos três estimadores para os parâmetros do modelo proposto. Ilustramos a utilidade do novo processo através de duas aplicações a conjuntos de dados reais.

Palavras-chave: Máxima verossimilhança condicional. processo INAR(1). distribuição série de potência.

Abstract

In this chapter, we introduce a first order non-negative integer valued autoregressive process with power series innovations based on the binomial thinning. This new model contains, as special cases, several models such as the Poisson INAR(1) model (Al-Osh and Alzaid, 1987), the geometric INAR(1) model (Jazi et al., 2012a) and many others. The main properties of the model are derived, such as mean, variance and the autocorrelation function. Yule-Walker, conditional least squares and conditional maximum likelihood estimators of the model parameters

are derived. An extensive Monte Carlo experiment is conducted to evaluate the performances of these estimators in finite samples. Special models are studied in some details. Applications to two real data sets are given to show the flexibility and potentiality of the new model.

Keywords: Conditional maximum likelihood. INAR(1) process. Power series distribution.

2.1 Introduction

In the last three decades, there has been a growing interest in discrete-valued time series models and several models for stationary processes with discrete marginal distributions have been proposed. Al-Osh and Alzaid (1987) proposed the first-order non-negative integer valued autoregressive (INAR(1)) process. Weiß (2009) proposed new autoregressive models for time series of binomial counts. Zhang et al. (2010) introduced pth-order integer valued autoregressive processes with a signed generalized power series thinning operator. Nastić et al. (2012) considered an integer valued autoregressive model of order p with geometric marginal distributions, using the negative binomial thinning. In a very recent paper, Jazi et al. (2012b) introduced a new stationary first-order integer valued autoregressive process with zero inflated Poisson innovations.

In general, detailed studies have been conducted not only on the formulation of models but also on properties (Silva and Oliveira, 2004), estimation (Jung et al., 2005), tests (Jung and Tremayne, 2003) and asymptotic distributions of model estimators (Freeland and McCabe, 2005) for different discrete marginal distributions.

Much theoretical work has been concentrated on the use of the Poisson distribution as an integral feature of the model. However, the Poisson distribution is not always suitable for modelling, since its mean and variance are the same and this property may be unacceptable for real data. Furthermore, in many real-life situations there are series which do not contain zeros in a large period of time or are even permanently positive. In these situations, the Poisson distribution is also not suitable for modelling. Zero truncation has been considered by Jazi et al. (2012b), who have recently proposed a first-order integer valued AR model with zero inflated Poisson innovations. This has also been studied by Zhu (2012), who has recently proposed integer-valued GARCH models that are based upon the zero inflated Poisson distribution and the zero inflated negative binomial distribution.

There is, therefore, a need to introduce different integer-valued time series models to deal with different particular real situations, like overdispersion or zero-inflation. The idea of considering a distribution for the innovations such that the marginal distribution of the observations will satisfy a given property has been extensively discussed in Weiß (2008), where approaches on how to obtain, for example, the overdispersed negative binomial or generalized Poisson distribution are presented.

In this context, the main purpose of this chapter is to propose a new first order non-negative integer valued autoregressive process with power series (PS) innovations based on the binomial thinning operator (Steutel and Van Harn, 1979). The motivation for such a process arises from its potential in modelling and analyzing non-negative integer valued time series when there is an indication of equidispersion, overdispersion, underdispersion or truncated distributions. The use of innovations that come from the PS family of distributions has many advantages, since this family of distributions constitutes a flexible framework for statistical modelling of discrete data in several real-life situations (Johnson et al., 2005).

We consider a sequence of discrete independent and identically distributed (i.i.d.) random variables $\{\epsilon_t; t \in \mathbb{Z}\}$, the distribution of each ϵ_t being indexed by a parameter θ and defined by the probability mass function

$$\Pr(\epsilon_t = x) = \frac{a(x)\,\theta^x}{C(\theta)},\ x \in \mathcal{S}, \tag{2.1}$$

where the support $\mathcal{S}$ of ϵ_t is a subset of the non-negative integers, $a(x) \geq 0$ depends only on x and there is $s > 0$ such that $C(\theta) = \sum_{x=0}^{\infty} a(x)\,\theta^x$ is finite for all $\theta \in (0, s)$ (s can be ∞). Although we will always consider θ as a value in $(0, s)$, we will also assume that the power series for $C(\theta)$ converges, in fact, to a finite value for $\theta \in (-s, s)$. If this is the case, then, $C(\theta)$ has derivatives of all orders in $(-s, s)$ and those derivatives can be obtained by differentiating the power series term to term. Also, because $a(x) \geq 0$ for all x, $C(\theta)$ and all its derivatives will be positive in $(0, s)$. For more details on the PS class of distributions, see Noack (1950).

Table 2.1 provides the functions $a(x)$, $C(\theta)$ and the parameter θ corresponding to some special cases of PS distributions such as the Bernoulli, binomial (with n being the integer number of replicas), geometric, Poisson, negative binomial and logarithmic distributions. For the negative binomial, there may exist situations for which we will want r to be integer-valued. In this case, x can be regarded as the random number of failures until exactly r successes are recorded in a sequence of independent trials where the probability of failure is θ. When using the binomial distribution, the value of n may be known in advance or may be estimated; the same holds for the value of r when using the negative binomial distribution.

Table 2.1: Some distributions in the family (1).

Distribution	$a(x)$	$C(\theta)$	s	$\mathcal{S}$
1. Bernoulli	1	$1+\theta$	∞	$\{0, 1\}$
2. Binomial	$\binom{n}{x}$	$(1+\theta)^n$	∞	$\{0, 1, \ldots, n\}$
3. Geometric	1	$(1-\theta)^{-1}$	1	$\{0, 1, 2, \ldots\}$
4. Poisson	$x!^{-1}$	e^{θ}	∞	$\{0, 1, 2, \ldots\}$
5. Negative Binomial	$\frac{\Gamma(r+x)}{x!\Gamma(r)}$	$(1-\theta)^{-r}$	1	$\{0, 1, 2, \ldots\}$
6. Logarithmic	x^{-1}	$-\log(1-\theta)$	1	$\{1, 2, 3, \ldots\}$

The chapter is structured as follows. The PSINAR(1) (power series INAR(1)) model is formally defined in Section 2.2 and some of its basic properties are outlined. In Section 2.3, we propose estimation methods for the model parameters. Three special cases of the proposed model are studied in Section 2.4. In Section 2.5, we present some simulation results for the estimation methods. In Section 2.6, we provide applications to two real data sets. The chapter is concluded in Section 2.7.

2.2 The PSINAR(1) model

Let Y be a non-negative integer valued random variable and $\alpha \in [0,1]$. According to Steutel and Van Harn (1979), the binomial thinning operator " $\circ$ " is defined as follows

$$\alpha \circ Y = \sum_{j=1}^{Y} Z_j,$$

where $\{Z_j\}_{j=1}^{Y}$ are i.i.d. random variables, independent of Y, with $\Pr(Z_j = 1) = 1 - \Pr(Z_j = 0) = \alpha$, i.e., $\{Z_j\}_{j=1}^{Y}$ is an i.i.d. Bernoulli random sequence. Given Y, $\alpha \circ Y$ has a binomial distribution with parameters (Y, α). For an account of the properties of the binomial thinning operator, see Silva and Oliveira (2004). Based on this operator, the first-order non-negative integer valued autoregressive PS model can be defined as

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t,\ t \in \mathbb{Z}, \tag{2.2}$$

where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d integer valued random variables with probability mass function satisfying (2.1), ϵ_t and Y_{t-i} being independent for all $i \geq 1$. Since $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence with finite mean and variance, we conclude that this sequence is a second-order stationary process. Consequently, the process $\{Y_t\}_{t\in\mathbb{Z}}$ satisfying (2.2) is second-order stationary if $0 \leq \alpha < 1$ (Du and Li, 1991).

We can view a realization of $\{Y_t\}_{t\in\mathbb{Z}}$ as having two components (Freeland and McCabe, 2004): the survivors of elements of Y_{t-1}, each with probability α of survival, and the elements which entered the system in the interval $(t-1, t]$ (the innovation term $\{\epsilon_t\}_{t\in\mathbb{Z}}$).

The question of which distribution to use for the $\{\epsilon_t\}_{t\in\mathbb{Z}}$ sequence may be rather subjective and it may depend also on the specific situation we are dealing with. For example, in epidemiology, suppose that a researcher is monitoring the number of individuals in a given population that did not contract a specific disease, that is, suppose Y_t is the number of healthy individuals in the population at time t. Let α be the probability that a healthy individual remains healthy, that is, does not contract the disease, in the next instant of time. Suppose also that, with a very small probability, a sick individual may become cured of this disease, such that, in the next time, we will have no more than one cured individual in the population. Then, the evolution of Y_t may be described by (2.2), using a Bernoulli distribution for ϵ_t. Suppose now that the same researcher, wishing to observe the evolution of cure in a very specific group, prescribes a given medicine to n sick individuals of that group, such that some of the individuals taking that medicine may become cured. Now the evolution of Y_t may be described by (2.2), using a binomial distribution for ϵ_t. When treating a serious disease, we can consider that a given individual gets cured or that this same individual ultimately dies. Then, a way of monitoring the efficiency of a given treatment is to observe how many individuals get cured before an individual dies. The evolution of Y_t may then be described by (2.2), with a geometric distribution for ϵ_t. If we observe how many individuals get cured before r individuals die, then, we can use a negative binomial distribution for ϵ_t.

A reasonable choice for the distribution of ϵ_t may also follow from statistical considerations. If it seems reasonable that the mean and variance of the distribution of the observations are equal, then, a simple Poisson model may be adequate. If variance seems to be smaller than the mean, we must discard the geometric and negative binomial distributions. A hypothesis test may be used to decide between a geometric and a negative binomial distribution. Also, it may seem reasonable that the observations are necessarily positive, which means that a truncated distribution should be used (see subsection 2.4.3).

From (2.2), it follows that $\{Y_t\}_{t\in\mathbb{Z}}$ is a Markov process. The proofs of the next two propositions can be seen in the Appendix.

Proposition 1. *For fixed $n \in \mathbb{Z}^+$, the transition probabilities of this process are given by*

$$\Pr(Y_t = k|Y_{t-1} = l) = \begin{cases} \dfrac{1}{C(\theta)} \displaystyle\sum_{i=0}^{\min(l,k-n)} \tbinom{l}{i}\alpha^i(1-\alpha)^{l-i}\,\theta^{k-i}\,a(k-i), & \text{if } \mathcal{S} = \{n, n+1, n+2, \ldots\}, \\ \dfrac{1}{C(\theta)} \displaystyle\sum_{i=\max(0,k-n)}^{\min(l,k)} \tbinom{l}{i}\alpha^i(1-\alpha)^{l-i}\,\theta^{k-i}\,a(k-i), & \text{if } \mathcal{S} = \{0, 1, 2, \ldots\, n\}, \end{cases} \tag{2.3}$$

for all $k, l \in \mathbb{Z}^+$, where $(\cdot)$ is the standard combinatorial symbol.

Proposition 2. *The Markov process defined by the transition probabilities above admits a unique stationary distribution.*

The marginal probability function of $\{Y_t\}_{t\in\mathbb{Z}}$ is given by

$$\Pr(Y_t = k) = \begin{cases} \dfrac{1}{C(\theta)} \displaystyle\sum_{l=0}^{\infty} \sum_{i=0}^{\min(l,k-n)} \tbinom{l}{i}\alpha^i(1-\alpha)^{l-i}\theta^{k-i}\,a(k-i)\Pr(Y_{t-1} = l), & \text{if } \mathcal{S} = \{n, n+1, n+2, \ldots\}, \\ \dfrac{1}{C(\theta)} \displaystyle\sum_{l=0}^{\infty} \sum_{i=\max(0,k-n)}^{\min(l,k)} \tbinom{l}{i}\alpha^i(1-\alpha)^{l-i}\theta^{k-i}\,a(k-i)\Pr(Y_{t-1} = l), & \text{if } \mathcal{S} = \{0, 1, 2, \ldots\, n\}, \end{cases}$$

which is a mixture distribution. It is important to highlight that the support of the distribution of Y_t is *not* a finite set, even if $\mathcal{S}$ is. In fact, it is not difficult to see that, for all $k, l \in \mathbb{Z}^+$, there will exist a positive integer m such that $\Pr(Y_m = k|Y_0 = l) > 0$.

Although we know a unique stationary distribution exists, obtaining such as stationary distribution from the above equations is in general a difficult task. Alternatively, if Ψ is the probability generating function for this stationary distribution, $\Psi(u) = \mathrm{E}(u^{Y_t})$, it is not difficult to check that Ψ is that function satisfying $C(\theta)\Psi(u) = C(u\,\theta)\Psi(\alpha\,u + (1-\alpha))$, for all u, θ. However, this approach is still not easy. For that very simple situation where $C(\theta) = \mathrm{e}^{\theta}$, which corresponds to the classical Poisson INAR(1) model, we will readily obtain that the stationary distribution is Poisson, its expected value being $\theta/(1-\alpha)$. On the other hand, the general problem of obtaining the stationary distribution of the observations, given a particular $C(\theta)$, seems to be, for most situations, quite difficult.

The moments of the random variable $\{\epsilon_t\}_{t\in\mathbb{Z}}$ can be easily obtained from the probability generating function $\Psi_{\epsilon_t}(u) = C(u\,\theta)/C(\theta)$. The expected value is $\mathrm{E}(\epsilon_t) = \mu_\epsilon = \theta\, G'(\theta)$ and the variance is $\mathrm{Var}(\epsilon_t) = \sigma^2_\epsilon = \theta^2 G''(\theta) + \mu_\epsilon$, where $G(\theta) = \log[C(\theta)]$, $G'(\theta) = d\,G(\theta)/d\theta$ and $G''(\theta) = d^2G(\theta)/d\theta^2$. These well-known results can be found, for example, in Johnson *et al.* (2005).

The mean and variance of the process $\{Y_t\}_{t\in\mathbb{Z}}$ as defined in (2.2) are, respectively

$$\mathrm{E}(Y_t) = \mu = \frac{\theta\, G'(\theta)}{1-\alpha}$$

and

$$\mathrm{Var}(Y_t) = \sigma^2 = \frac{\theta}{1-\alpha}\left[G'(\theta) + \frac{\theta\, G''(\theta)}{1+\alpha}\right] = \mu + \frac{\theta^2 G''(\theta)}{1-\alpha^2}.$$

Note that the variance can be smaller or greater than the mean, depending on the sign of $G''(\theta)$. The dispersion index, which is the variance to mean ratio, is given by

$$\frac{\sigma^2}{\mu} = 1 + \frac{\theta\, G''(\theta)}{(1+\alpha)G'(\theta)}.$$

Also, the mean and variance will be equal when G is a linear function, which is the case for the Poisson distribution.

The expressions for the moments of the conditional and unconditional distributions of the observations in a general INAR(1) process can be found, for example, in Rajarshi (2012). For our specific process, we obtain the conditional expectation as

$$\mathrm{E}(Y_t|Y_{t-1}) = \alpha\, Y_{t-1} + \theta\, G'(\theta)$$

and the conditional variance as

$$\mathrm{Var}(Y_t|Y_{t-1}) = \alpha(1-\alpha)Y_{t-1} + \theta\, G'(\theta) + \theta^2 G''(\theta).$$

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

$$\mathrm{Corr}(Y_t, Y_{t-k}) = \rho(k) = \alpha^k,\ k \geq 1, \tag{2.4}$$

which obviously is restricted to be positive.

Next, we consider the problem of estimating the parameters.

2.3 Estimation of the unknown parameters

This section is concerned with the estimation of the two parameters of interest. We consider three estimation methods, namely, Yule-Walker, conditional least squares and conditional maximum likelihood.

2.3.1 Yule-Walker estimation

From a sample $Y_1, \ldots, Y_T$ of a stationary process $\{Y_t\}_{t\in\mathbb{Z}}$, the sample autocorrelation function is given by

$$\widehat{\rho}(k) = \frac{\sum\limits_{t=1}^{T-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2},$$

where $\overline{Y} = (1/T)\sum_{t=1}^{T} Y_t$ is the sample mean. The Yule-Walker (YW) estimator of α, based upon the fact that $\rho(k) = \alpha^k$, as in (2.4), is given by

$$\widehat{\alpha} = \widehat{\rho}(1) = \frac{\sum\limits_{t=1}^{T-1}(Y_t - \overline{Y})(Y_{t+1} - \overline{Y})}{\sum_{t=1}^{T}(Y_t - \overline{Y})^2}. \tag{2.5}$$

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

$$\widehat{\mu}_\epsilon = (1-\widehat{\alpha})\overline{Y},$$

where $\widehat{\alpha}$ is given in (2.5). An estimator of the parameter θ can be obtained as the solution of the equation $\widehat{\theta}\,G'(\widehat{\theta}) = (1-\widehat{\alpha})\overline{Y}$. The estimator of θ may have closed-form, depending on which distribution is being used.

Du and Li (1991) demonstrated that the usual mean estimator, the autocovariance and autocorrelation functions, given by $\overline{Y}$, $\widehat{\gamma}(k) = (1/T)\sum_{t=1}^{T-k}(Y_t - \overline{Y})(Y_{t-k} - \overline{Y})$ and $\widehat{\rho}(k) = \widehat{\gamma}(k)/\widehat{\gamma}(0)$, $0 \leq k \leq T-1$, respectively, are strongly consistent.

2.3.2 Conditional least squares estimation

The conditional least squares estimator $\widehat{\boldsymbol{\eta}} = (\widehat{\alpha}, \widehat{\mu}_\epsilon)^\top$ of $\boldsymbol{\eta} = (\alpha, \mu_\epsilon)^\top$ is given by

$$\widehat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \left\{S_T(\boldsymbol{\eta})\right\},$$

where $S_T(\boldsymbol{\eta}) = \sum_{t=2}^{T}[Y_t - g(\boldsymbol{\eta}, Y_{t-1})]^2$ and $g(\boldsymbol{\eta}, Y_{t-1}) = \mathrm{E}(Y_t|Y_{t-1})$. Thus, following Klimko and Nelson (1978), the conditional least squares (CLS) estimators of α and μ_ϵ can be expressed in closed-form as

$$\widehat{\alpha} = \frac{\sum\limits_{t=2}^{T} Y_t Y_{t-1} - \dfrac{1}{T-1}\sum\limits_{t=2}^{T} Y_t \sum\limits_{t=2}^{T} Y_{t-1}}{\sum\limits_{t=2}^{T} Y_{t-1}^2 - \dfrac{1}{T-1}\left(\sum\limits_{t=2}^{T} Y_{t-1}\right)^2} \tag{2.6}$$

and

$$\widehat{\mu}_\epsilon = \frac{1}{T-1}\left(\sum_{t=2}^{T} Y_t - \widehat{\alpha}\sum_{t=2}^{T} Y_{t-1}\right),$$

where $\widehat{\alpha}$ is given in (2.6). As in Section 2.3.1, the estimator of the parameter θ can be obtained by solving the equation $\widehat{\theta}\,G'(\widehat{\theta}) = \widehat{\mu}_\epsilon$. The estimator of θ may have closed-form, depending on which distribution is being used.

2.3.3 Conditional maximum likelihood estimation

Suppose that y_1 is fixed. The conditional log-likelihood function for the PSINAR(1) model is given by

$$\ell(\alpha,\theta) = \log\left(\prod_{t=2}^{T} \Pr(Y_t|Y_{t-1})\right) = \sum_{t=2}^{T} \log(\Pr(Y_t|Y_{t-1})), \tag{2.7}$$

where $\Pr(Y_t|Y_{t-1})$ is given by (2.3).

The conditional maximum likelihood (CML) estimators $\widehat{\alpha}$ and $\widehat{\theta}$ of α and θ are defined as the values of α and θ that maximize the conditional log-likelihood function in (2.7). There will be, in general, no closed-form for the CML estimates and to obtain them, we will need, in practice, numerical methods.

2.4 Special cases

In this section, we investigate some special cases of the PSINAR(1) model and provide expressions for the mean and variance.

2.4.1 Geometric INAR(1) model

For $\mathcal{S} = \mathbb{Z}^+$ and $C(\theta) = (1-\theta)^{-1}, \theta \in (0,1)$ in (2.1), we say that $\{Y_t\}_{t\in\mathbb{Z}}$ is a geometric INAR(1) model. Alzaid and Al-Osh (1988) introduced the INAR(1) process with geometric marginal distribution. Ristić et al. (2009) proposed the first-order integer valued autoregressive process with geometric marginal distribution based on negative binomial thinning. Jazi et al. (2012a) studied the geometric INAR(1) process.

The transition probabilities of this process are given by

$$\Pr(Y_t = k|Y_{t-1} = l) = (1-\theta) \sum_{i=0}^{\min(k,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i}\, \theta^{k-i}, \;\; 0 < \theta < 1.$$

The mean and variance of $\{Y_t\}_{t\in\mathbb{Z}}$ are

$$\mathrm{E}(Y_t) = \mu = \frac{\theta}{(1-\alpha)(1-\theta)} \;\text{ and }\; \mathrm{Var}(Y_t) = \sigma^2 = \frac{\theta + \alpha\,\theta(1-\theta)}{(1-\alpha^2)(1-\theta)^2}.$$

The conditional expectation and the conditional variance are given, respectively, by

$$\mathrm{E}(Y_t|Y_{t-1}) = \alpha\, Y_{t-1} + \frac{\theta}{1-\theta}$$

and

$$\mathrm{Var}(Y_t|Y_{t-1}) = \alpha(1-\alpha)Y_{t-1} + \frac{\theta}{(1-\theta)^2}.$$

Note that $\mu = (1-\alpha)^{-1}[(1-\theta)^{-1} - 1]$ is an increasing function of α and θ. Also,

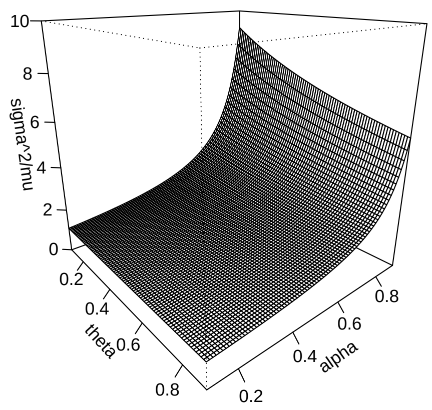
$$\sigma^2 = \frac{1}{1-\alpha^2}\frac{\theta}{(1-\theta)^2} + \frac{\alpha}{1-\alpha^2}\frac{\theta}{1-\theta}$$

is an increasing function of α and θ. Furthermore, we can easily obtain

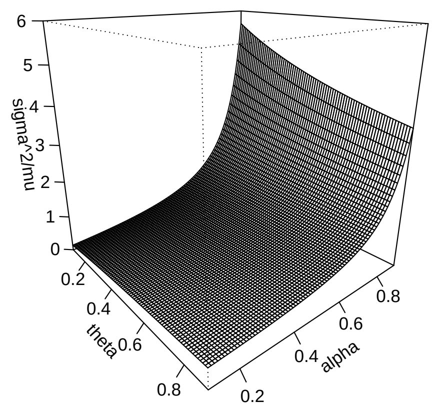
$$\frac{\sigma^2}{\mu} = 1 + \frac{\theta}{(1+\alpha)(1-\theta)} > 1.$$

The geometric INAR(1) process, therefore, may be used as a model for overdispersed non-negative integer valued time series. From the above equation, we can readily conclude that the variance-mean ratio is an increasing function of θ, but a decreasing function of α.

Figure 2.1(a) shows how σ^2/μ behaves as a function of α and θ. For more details about the geometric INAR(1) process, see Jazi et al. (2012a).



(a) Geometric INAR(1)



(b) Logarithmic INAR(1)

Figure 2.1: Plot of the variance-to-mean ratio against α and θ.

2.4.2 Poisson INAR(1) model

For $\mathcal{S} = \mathbb{Z}^+$ and $C(\theta) = \mathrm{e}^\theta$ in (2.1), we say that $\{Y_t\}_{t\in\mathbb{Z}}$ is a Poisson INAR(1) model. Al-Osh and Alzaid (1987) proposed and studied the Poisson INAR(1) process. Many new results on it have been obtained in recent years. For example, Hellström (2001) focused on unit root testing, Freeland and McCabe (2005) obtained asymptotic properties of CLS estimators, Weiß (2011) proposed several asymptotic simultaneous confidence regions for the two parameters.

The transition probabilities of this process are given by

$$\Pr(Y_t = k | Y_{t-1} = l) = \mathrm{e}^{-\theta} \sum_{i=0}^{\min(k,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \frac{\theta^{k-i}}{(k-i)!}, \ \ \theta > 0.$$

The mean and variance of $\{Y_t\}_{t\in\mathbb{Z}}$ are

$$\mathrm{E}(Y_t) = \mu = \frac{\theta}{1-\alpha} \ \text{ and } \ \mathrm{Var}(Y_t) = \sigma^2 = \frac{\theta}{1-\alpha}.$$

The mean and variance are equal for this model, both increasing with θ and α. For more details about the Poisson INAR(1) process, see Al-Osh and Alzaid (1987).

2.4.3 Truncated models

Truncated Poisson and negative binomial models have been discussed, among others, by Creel and Loomis (1990) and Grogger and Carson (1991).

The PSINAR(1) model defined by (2.2) has the flexibility of modelling data by a Markovian process for which the state space is some proper subset of the nonnegative integers. This can be achieved, for example, by considering truncated distributions for the innnovations.

Table 2.2 provides the functions $a(x)$ and $C(\theta)$ corresponding to some special cases of the PS distributions truncated at zero.

Table 2.2: Some distributions truncated at zero in the family (1).

Distribution	$a(x)$	$C(\theta)$	s	$\mathcal{S}$
1. Binomial	$\binom{n}{x}$	$(1+\theta)^n - 1$	∞	$\{1,2\ldots,n\}$
2. Geometric	1	$\theta(1-\theta)^{-1}$	1	$\{1,2,3,\ldots\}$
3. Poisson	$x!^{-1}$	$e^{\theta} - 1$	∞	$\{1,2,3,\ldots\}$

2.4.4 Logarithmic INAR(1) model

The logarithmic INAR(1) can also be a model for a series of counts where zeros are not observed. For $\mathcal{S} = \{1,2,3,\ldots\}$ and $C(\theta) = -\log(1-\theta), \theta \in (0,1)$ in (2.1), we define $\{Y_t\}_{t\in\mathbb{Z}}$ as the logarithmic INAR(1) model. The transition probabilities for this process are given by

$$\Pr(Y_t = k|Y_{t-1} = l) = -\frac{1}{\log(1-\theta)} \sum_{i=0}^{\min(k-1,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \frac{\theta^{k-i}}{k-i}.$$

The mean and variance of $\{Y_t\}_{t\in\mathbb{Z}}$ are

$$\mathrm{E}(Y_t) = \mu = \frac{a\,\theta}{(1-\alpha)(1-\theta)}$$

and

$$\mathrm{Var}(Y_t) = \sigma^2 = \frac{a\,\theta[\alpha(1-\theta) + (1-a\,\theta)]}{(1-\alpha^2)(1-\theta)^2} = \frac{\mu[\alpha(1-\theta) + (1-a\,\theta)]}{(1+\alpha)(1-\theta)},$$

where $a = -1/\log(1-\theta)$. The conditional expectation and the conditional variance are given by

$$\mathrm{E}(Y_t|Y_{t-1}) = \alpha\, Y_{t-1} + \frac{a\,\theta}{1-\theta}$$

and

$$\mathrm{Var}(Y_t|Y_{t-1}) = \alpha(1-\alpha)Y_{t-1} + \frac{a\,\theta(1-a\,\theta)}{(1-\theta)^2}.$$

Both mean and variance are increasing functions of α and θ. The variance-mean ratio can be easily seen to be

$$\frac{\sigma^2}{\mu} = 1 + \frac{\theta(1-a)}{(1+\alpha)(1-\theta)};$$

it follows that there is

- equidispersion when $\theta = 1 - \mathrm{e}^{-1}$,
- overdispersion when $\theta > 1 - \mathrm{e}^{-1}$,
- underdispersion when $\theta < 1 - \mathrm{e}^{-1}$.

Figure 2.1(b) shows how σ^2/μ behaves as a function of α and θ.

2.5 Monte Carlo simulation study

The performances of the YW, CLS and CML estimators for a sample of T observed values of $\{Y_t\}$ is the motivation of this section. Some numerical results for different values of the parameters α and θ are presented in Tables 2.3, 2.4 and 2.5. The sample sizes considered were $T = 100, 200$ and 300. The Monte Carlo simulation experiments were performed using the R programming language; see `http://www.r-project.org`. The number of Monte Carlo replications was 1000. The CML estimates of α and θ are obtained by maximizing the conditional log-likelihood function using the BFGS quasi-Newton nonlinear optimization algorithm with numerical derivatives. As starting values for the algorithm, we suggest the estimates obtained by Yule-Walker method. For each different situation, we have estimated the bias and the mean squared error (MSE). The YW and CLS estimates for θ are not obtained directly for the logarithmic INAR(1) model. In this case, the estimate of θ is obtained as the value of θ that minimizes $g(\theta) = \{\theta/[(1-\theta)\log(1-\theta)] - \widehat{\mu}_\epsilon\}^2$.

Tables 2.3, 2.4 and 2.5 present the biases and MSE's (given in parentheses) of the different estimators for geometric INAR(1), Poisson INAR(1) and logarithmic INAR(1) models, respectively. It is noteworthy that the CML estimators of the parameters α and θ display biases and MSE's that are much smaller than those of the corresponding YW and CLS for almost all sample sizes considered in the experiment. Note that as the sample size increases, the bias tends to zero in all three cases, confirming that the estimators are asymptotically unbiased.

It is expected for the CML estimator to display the best performance, since it uses the whole information of the distribution. The empirical investigation presented here suggests that, generally speaking, the CML is, in fact, much better than the YW and CLS. Thus, we recommend the use of the CML method to estimate the model parameters of an INAR(1) process with PS innovation.

Table 2.3: Bias and MSE (in parentheses) of the parameters in geometric INAR(1).

Sample size	Parameters	CLS	YW	CML
$\alpha = 0.3, \theta = 0.3$				
$T = 100$	$\widehat{\alpha}$	−0.0216 (0.0120)	−0.0245 (0.0119)	−0.0097 (0.0078)
	$\widehat{\theta}$	0.0044 (0.0023)	0.0054 (0.0023)	0.0013 (0.0019)
$T = 200$	$\widehat{\alpha}$	−0.0086 (0.0064)	−0.0101 (0.0064)	−0.0027 (0.0038)
	$\widehat{\theta}$	0.0011 (0.0012)	0.0017 (0.0012)	−0.0003 (0.0009)
$T = 300$	$\widehat{\alpha}$	−0.0056 (0.0042)	−0.0065 (0.0042)	−0.0023 (0.0026)
	$\widehat{\theta}$	0.0002 (0.0008)	0.0005 (0.0008)	−0.0001 (0.0006)
$\alpha = 0.7, \theta = 0.3$				
$T = 100$	$\widehat{\alpha}$	−0.0362 (0.0079)	−0.0435 (0.0085)	−0.0065 (0.0025)
	$\widehat{\theta}$	0.0166 (0.0042)	0.0215 (0.0043)	−0.0009 (0.0020)
$T = 200$	$\widehat{\alpha}$	−0.0168 (0.0036)	−0.0204 (0.0037)	−0.0042 (0.0012)
	$\widehat{\theta}$	0.0079 (0.0020)	0.0104 (0.0020)	0.0007 (0.0010)
$T = 300$	$\widehat{\alpha}$	−0.0100 (0.0024)	−0.0123 (0.0024)	−0.0007 (0.0008)
	$\widehat{\theta}$	0.0052 (0.0015)	0.0068 (0.0015)	−0.0005 (0.0006)
$\alpha = 0.3, \theta = 0.7$				
$T = 100$	$\widehat{\alpha}$	−0.0232 (0.0097)	−0.0262 (0.0097)	0.0003 (0.0027)
	$\widehat{\theta}$	0.0024 (0.0013)	0.0033 (0.0013)	−0.0027 (0.0008)
$T = 200$	$\widehat{\alpha}$	−0.0095 (0.0048)	−0.0108 (0.0047)	0.0024 (0.0014)
	$\widehat{\theta}$	−0.0022 (0.0008)	−0.0020 (0.0008)	−0.0020 (0.0005)
$T = 300$	$\widehat{\alpha}$	−0.0077 (0.0033)	−0.0087 (0.0033)	−0.0005 (0.0009)
	$\widehat{\theta}$	0.0011 (0.0005)	0.0014 (0.0005)	−0.0008 (0.0003)
$\alpha = 0.7, \theta = 0.7$				
$T = 100$	$\widehat{\alpha}$	−0.0335 (0.0070)	−0.0411 (0.0076)	−0.0016 (0.0009)
	$\widehat{\theta}$	0.0121 (0.0030)	0.0169 (0.0029)	−0.0024 (0.0010)
$T = 200$	$\widehat{\alpha}$	−0.0163 (0.0035)	−0.0195 (0.0036)	−0.0003 (0.0004)
	$\widehat{\theta}$	0.0053 (0.0017)	0.0073 (0.0017)	−0.0014 (0.0005)
$T = 300$	$\widehat{\alpha}$	−0.0099 (0.0019)	−0.0122 (0.0019)	0.0009 (0.0002)
	$\widehat{\theta}$	0.0042 (0.0010)	0.0056 (0.0010)	−0.0009 (0.0002)

Table 2.4: Bias and MSE (in parentheses) of the parameters in Poisson INAR(1).

Sample size	Parameters	CLS	YW	CML
$\alpha = 0.3, \theta = 1.0$				
$T = 100$	$\widehat{\alpha}$	−0.0236 (0.0106)	−0.0265 (0.0105)	−0.0133 (0.0087)
	$\widehat{\theta}$	0.0303 (0.0282)	0.0348 (0.0284)	0.0160 (0.0251)
$T = 200$	$\widehat{\alpha}$	−0.0094 (0.0052)	−0.0109 (0.0052)	−0.0061 (0.0043)
	$\widehat{\theta}$	0.0119 (0.0146)	0.0143 (0.0145)	0.0072 (0.0128)
$T = 300$	$\widehat{\alpha}$	−0.0063 (0.0035)	−0.0073 (0.0035)	−0.0022 (0.0028)
	$\widehat{\theta}$	0.0098 (0.0096)	0.0114 (0.0096)	0.0038 (0.0080)
$\alpha = 0.7, \theta = 1.0$				
$T = 100$	$\widehat{\alpha}$	−0.0317 (0.0068)	−0.0389 (0.0072)	−0.0050 (0.0022)
	$\widehat{\theta}$	0.0926 (0.0800)	0.1161 (0.0852)	0.0037 (0.0276)
$T = 200$	$\widehat{\alpha}$	−0.0155 (0.0033)	−0.0189 (0.0034)	−0.0024 (0.0011)
	$\widehat{\theta}$	0.0502 (0.0367)	0.0613 (0.0378)	0.0070 (0.0136)
$T = 300$	$\widehat{\alpha}$	−0.0090 (0.0019)	−0.0113 (0.0020)	−0.0004 (0.0007)
	$\widehat{\theta}$	0.0335 (0.0236)	0.0410 (0.0242)	0.0045 (0.0089)
$\alpha = 0.3, \theta = 2.0$				
$T = 100$	$\widehat{\alpha}$	−0.0160 (0.0103)	−0.0187 (0.0102)	−0.0056 (0.0094)
	$\widehat{\theta}$	0.0445 (0.0965)	0.0531 (0.0960)	0.0147 (0.0886)
$T = 200$	$\widehat{\alpha}$	−0.0074 (0.0049)	−0.0088 (0.0049)	−0.0028 (0.0041)
	$\widehat{\theta}$	0.0194 (0.0482)	0.0236 (0.0481)	0.0065 (0.0418)
$T = 300$	$\widehat{\alpha}$	−0.0061 (0.0034)	−0.0072 (0.0034)	−0.0013 (0.0028)
	$\widehat{\theta}$	0.0178 (0.0312)	0.0207 (0.0312)	0.0038 (0.0266)
$\alpha = 0.7, \theta = 2.0$				
$T = 100$	$\widehat{\alpha}$	−0.0335 (0.0075)	−0.0405 (0.0079)	−0.0029 (0.0022)
	$\widehat{\theta}$	0.2270 (0.3462)	0.2721 (0.3648)	0.0219 (0.1024)
$T = 200$	$\widehat{\alpha}$	−0.0164 (0.0033)	−0.0199 (0.0034)	−0.0026 (0.0011)
	$\widehat{\theta}$	0.1064 (0.1516)	0.1299 (0.1569)	0.0149 (0.0533)
$T = 300$	$\widehat{\alpha}$	−0.0111 (0.0021)	−0.0134 (0.0021)	−0.0009 (0.0007)
	$\widehat{\theta}$	0.0708 (0.0926)	0.0867 (0.0950)	0.0023 (0.0306)

Table 2.5: Bias and MSE (in parentheses) of the parameters in logarithmic INAR(1).

Sample size	Parameters	CLS	YW	CML
$\alpha = 0.3, \theta = 0.3$				
$T = 100$	$\widehat{\alpha}$	−0.0171 (0.0112)	−0.0198 (0.0111)	−0.0007 (0.0026)
	$\widehat{\theta}$	−0.0211 (0.0562)	0.0147 (0.0537)	−0.0105 (0.0075)
$T = 200$	$\widehat{\alpha}$	−0.0100 (0.0061)	−0.0116 (0.0061)	−0.0010 (0.0013)
	$\widehat{\theta}$	−0.0085 (0.0250)	−0.0052 (0.0245)	−0.0064 (0.0040)
$T = 300$	$\widehat{\alpha}$	−0.0079 (0.0040)	−0.0088 (0.0039)	−0.0001 (0.0008)
	$\widehat{\theta}$	−0.0044 (0.0149)	−0.0025 (0.0131)	−0.0040 (0.0026)
$\alpha = 0.7, \theta = 0.3$				
$T = 100$	$\widehat{\alpha}$	−0.0340 (0.0076)	−0.0408 (0.0081)	−0.0002 (0.0008)
	$\widehat{\theta}$	0.0169 (0.1167)	0.0466 (0.1039)	−0.0126 (0.0072)
$T = 200$	$\widehat{\alpha}$	−0.0160 (0.0033)	−0.0195 (0.0035)	−0.0001 (0.0004)
	$\widehat{\theta}$	0.0059 (0.0588)	0.0212 (0.0553)	−0.0066 (0.0035)
$T = 300$	$\widehat{\alpha}$	−0.0134 (0.0021)	−0.0156 (0.0022)	0.0001 (0.0002)
	$\widehat{\theta}$	0.0032 (0.0343)	0.0080 (0.0331)	−0.0046 (0.0024)
$\alpha = 0.3, \theta = 0.7$				
$T = 100$	$\widehat{\alpha}$	−0.0219 (0.0113)	−0.0251 (0.0113)	−0.0006 (0.0021)
	$\widehat{\theta}$	−0.0149 (0.0154)	−0.0119 (0.0171)	−0.0091 (0.0029)
$T = 200$	$\widehat{\alpha}$	−0.0104 (0.0052)	−0.0119 (0.0052)	0.0001 (0.0010)
	$\widehat{\theta}$	−0.0054 (0.0040)	−0.0042 (0.0039)	−0.0045 (0.0014)
$T = 300$	$\widehat{\alpha}$	−0.0068 (0.0035)	−0.0078 (0.0035)	−0.0001 (0.0007)
	$\widehat{\theta}$	−0.0033 (0.0024)	−0.0026 (0.0024)	−0.0022 (0.0009)
$\alpha = 0.7, \theta = 0.7$				
$T = 100$	$\widehat{\alpha}$	−0.0354 (0.0081)	−0.0433 (0.0088)	−0.0004 (0.0006)
	$\widehat{\theta}$	−0.0276 (0.1968)	−0.0103 (0.1397)	−0.0072 (0.0028)
$T = 200$	$\widehat{\alpha}$	−0.0143 (0.0031)	−0.0180 (0.0032)	−0.0002 (0.0003)
	$\widehat{\theta}$	−0.0050 (0.0380)	0.0088 (0.0371)	−0.0035 (0.0014)
$T = 300$	$\widehat{\alpha}$	−0.0087 (0.0020)	−0.0109 (0.0020)	0.0001 (0.0002)
	$\widehat{\theta}$	0.0016 (0.0100)	−0.0047 (0.0122)	−0.0020 (0.0010)

2.6 Applications to real data

We assess the usefulness of the proposed model in an analysis of real data. The first data set is obtained from the crime data section of the forecasting principles site (http://www.forecastingprinciples.-com). This data series represents the counting of sex offences reported in the 21st police car beat in Pittsburgh, during one month. The data consist of 144 observations, starting in January 1990 and ending in December 2001. These data were previously studied by Ristić et al. (2009) and are listed in Table 2.6. The required numerical evaluations were carried out using the R software.

Table 2.6: Sex offences.

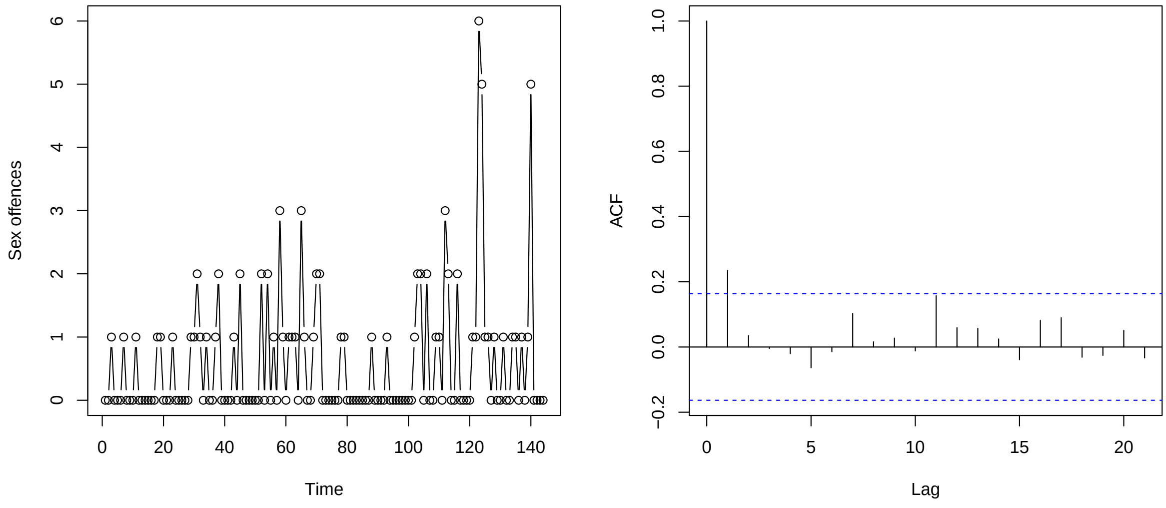
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1990	0	0	1	0	0	0	1	0	0	0	1	0
1991	0	0	0	0	0	1	1	0	0	0	1	0
1992	0	0	0	0	1	1	2	1	0	1	0	0
1993	1	2	0	0	0	0	1	0	2	0	0	0
1994	0	0	0	2	0	2	0	1	0	3	1	0
1995	1	1	1	0	3	1	0	0	1	2	2	0
1996	0	0	0	0	0	1	1	0	0	0	0	0
1997	0	0	0	1	0	0	0	0	1	0	0	0
1998	0	0	0	0	0	1	2	2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

Table 2.7 provides some descriptive statistics. Note that the data set includes the value 0. Thus, the logarithmic INAR(1) model is not appropriate. Furthermore, the sample variance is much larger than the sample mean, and then, the data seem to be overdispersed. Consequently, a Poisson marginal distribution would not be appropriate. The series and its sample autocorrelation are displayed in Figure 2.2.

Table 2.7: Descriptive statistics.

Min.	Q_1	Q_2	Mean	$\widehat{\rho}(1)$	Q_3	Max.	Var.
0.0000	0.0000	0.0000	0.5903	0.2348	1.0000	6.0000	1.0268

Analyzing Figure 2.2 we conclude that first order autoregressive models may be appropriate for the given data series. The behavior of the series indicates that it may be a mean stationary time series. We compared the geometric INAR(1) with the negative binomial INAR(1) (corresponding to $C(\theta) = (1-\theta)^{-r}$) and also with the geometric first-order integer valued autoregressive (NGINAR(1)) model with geometric marginal distribution (Ristić et al., 2009). Table 2.8 provides the CML estimates (with corresponding standard errors in parentheses) of the model parameters and three goodness-of-fit statistics: AIC (Akaike information criterion), RMS (root mean square of differences between observations and predicted values) and MA (absolute mean of differences between observations and predicted



(a) Series

(b) Sample autocorrelation

Figure 2.2: Counting of sex offences in Pittsburgh with sample ACF.

values). Since the expected information matrix is not available, the standard errors are obtained as the square roots of the elements in the diagonal of the inverse of the negative of the Hessian of the conditional log-likelihood calculated at the CML estimates.

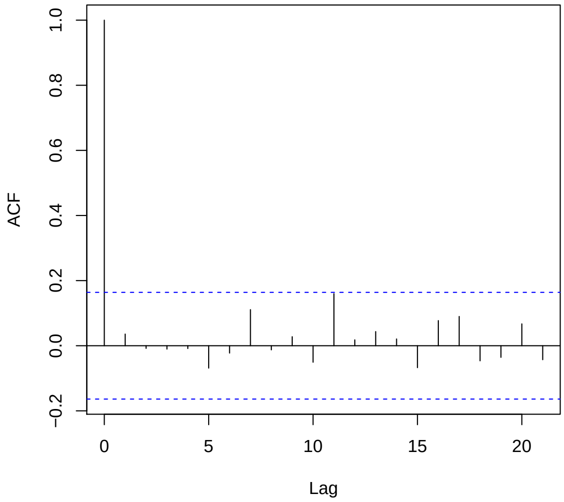
Table 2.8: Estimated parameters (with corresponding standard errors in parentheses), AIC, RMS and MA.

Model	CML estimates	AIC	RMS	MA
Geometric INAR(1)	$\widehat{\alpha} = 0.1143$ (0.0754) $\widehat{\theta} = 0.3449$ (0.0364)	302.57	0.9913	0.7270
Negative Binomial INAR(1)	$\widehat{\alpha} = 0.2021$ (0.0660) $\widehat{\theta} = 0.0794$ (0.0103) $\widehat{r} = 5.4993$ (0.0001)	102.15	0.9842	0.7237
NGINAR(1)	$\widehat{\alpha} = 0.1660$ (0.0965) $\widehat{\mu} = 0.5929$ (0.0958)	301.75	0.9862	0.7235

From Table 2.8, we conclude that the geometric INAR(1) and the NGINAR(1) models are competitive. Also, we can compare the AIC's to conclude that the proposed negative binomial INAR(1) model produces much better fits to the data. The estimated model is

$$Y_t = 0.2021 \circ Y_{t-1} + \epsilon_t,$$

where $\epsilon_t \sim$ Negative Binomial$(5.50, 0.92)$. The sample autocorrelations of the residuals, $\widehat{\epsilon}_t$, are displayed in Figure 2.3, where $\widehat{\epsilon}_t = Y_t - \widehat{\alpha}\, Y_{t-1} - \widehat{\lambda}$.



(a) Sample autocorrelation of the residuals

Figure 2.3: Sample autocorrelations of the residuals obtained from Negative Binomial INAR(1) model.

The second data set is given by Bakouch and Ristić (2010) as an application of their ZTPINAR(1) model, for which the marginal distribution of the observations is a zero truncated Poisson. Their original data counts family violences in the 11th police car beat in Pittsburgh during one month. The data set is obtained from the crime data section of the forecasting principles site (http://www.forecastingprinciples.com). It consists of 144 observations, starting in January 1990 and ending in December 2001. In order to use their zero truncated Poisson model, the authors transformed the series, adding 1 to each observation. The transformed data are listed in Table 2.9.

Table 2.10 provides some descriptive statistics. We see that the transformed data set does not assume the value 0. Thus, the logarithmic INAR(1) and truncated Poisson INAR(1) may be also appropriate. The transformed series and its sample autocorrelations are displayed in Figure 2.4.

Analyzing Figure 2.4 we conclude that the first order autoregressive models may be appropriate for the current data series. The behavior of the series indicates that it may be a mean stationary time series. We compared the logarithmic INAR(1) and the truncated Poisson INAR(1) (corresponding to $C(\theta) = e^{\theta} - 1$) fittings with that of the ZTPINAR(1).

Table 2.11 provides the CML estimates (with corresponding standard errors in parentheses) of the model parameters and goodness-of-fit statistics. From this table, we observe that the three models are competitive, the first two being only marginally better. The estimated truncated Poisson INAR(1) model, which is only marginally better than the other two, is

$$Y_t = 0.2145 \circ Y_{t-1} + \epsilon_t,$$

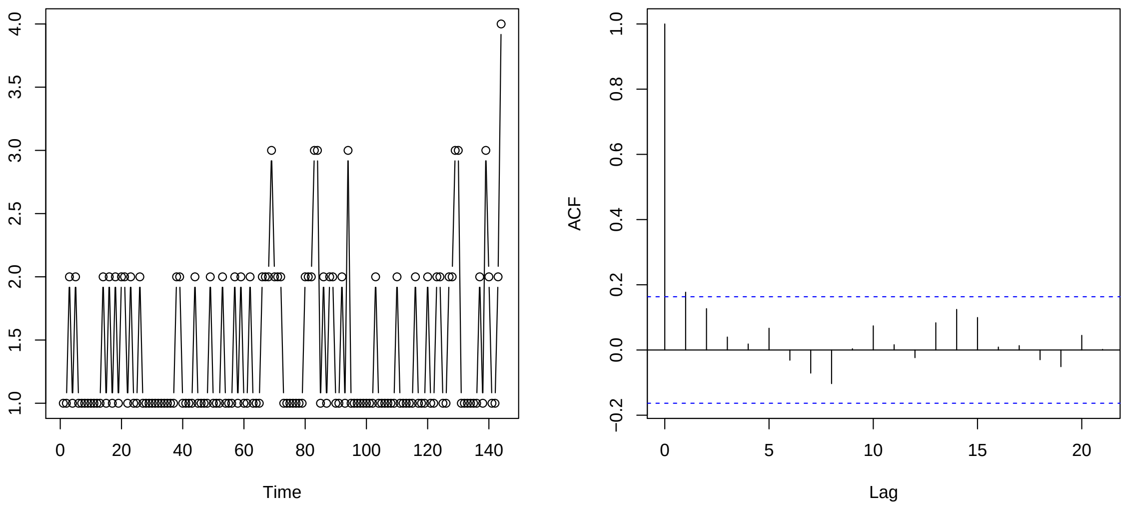
where $\epsilon_t \sim$ Truncated Poisson(0.2356). The sample autocorrelations of the residuals are shown in Figure 2.5 .

Table 2.9: Family violences.

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

Table 2.10: Descriptive statistics.

Min.	Q_1	Q_2	Mean	$\widehat{\rho}(1)$	Q_3	Max.	Var.
1.0000	1.0000	1.000	1.403	0.1770	2.0000	4.0000	0.3821



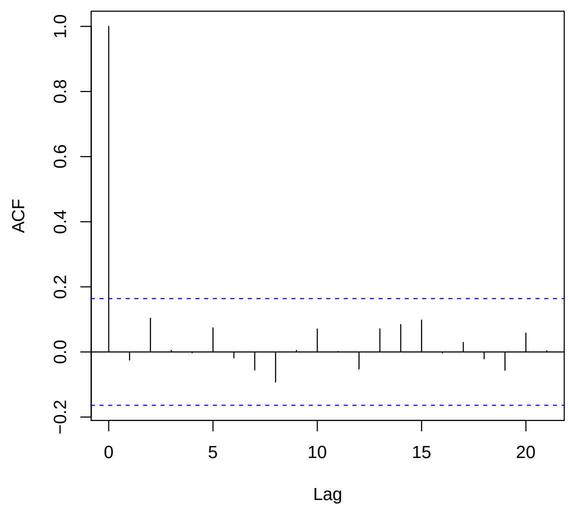
(a) Serie

(b) Sample autocorrelation

Figure 2.4: The sample path of the second time series and the autocorrelation function.

Table 2.11: Estimated parameters (with corresponding standard errors in parentheses), AIC, RMS and MA.

Model	CML estimates	AIC	RMS	MA
Logarithmic INAR(1)	$\widehat{\alpha} = 0.2199\ (0.0447)$ $\widehat{\theta} = 0.1727\ (0.0798)$	233.21	0.6061	0.5205
Truncated Poisson INAR(1)	$\widehat{\alpha} = 0.2045\ (0.0569)$ $\widehat{\theta} = 0.2356\ (0.1378)$	232.87	0.6059	0.5214
ZTPINAR(1)	$\widehat{\alpha} = 0.4202\ (0.0878)$ $\widehat{\lambda} = 0.7450\ (0.1142)$	233.46	0.6061	0.5221



(a) Sample autocorrelations of the residuals

Figure 2.5: Sample autocorrelations of the residuals obtained from truncated Poisson INAR(1) model.

2.7 Concluding remarks

In this chapter, we introduced first order non-negative integer valued autoregressive processes with power series innovations based on binomial thinning. The main properties of the model are derived, such as the mean, variance, autocorrelation function and transition probabilities. Three methods for estimating the model parameters are considered. Special models (Geometric INAR(1), Poisson INAR(1) and Logarithmic INAR(1) models) are studied in some detail. We observe that the use of innovations that come from a PS distribution has many advantages, and allows us to create processes for modelling series of counts in several real-life situations. Indeed, the general Markovian process that is studied in this paper offers many modelling options to the user depending on the characteristics of the data. In the simulation study, we compare YW, CLS and CML estimators. The simulation results indicate that the YW and CLS methods produce estimators with similar performances and that the CML is much better. Thus, we recommend the use of the CML method to estimate the model parameters of an INAR(1) process with PS innovations. Finally, we fitted PS models to two real data sets to show the potentiality of the new model. These applications also demonstrate the practical relevance of the new model.

2.8 Appendix

Proof of Proposition 1

For the PSINAR(1) process with binomial thinning operation, the conditional distribution of Y_t given Y_{t-1} is the convolution of the binomial distribution of the result of the thinning operation, $\alpha \circ Y_{t-1}$, with the PS distribution of the innovation process, ϵ_t (Sprott, 1983). Thus, let $\bullet$ denote convolution, let

$$f_1(i) = \binom{Y_{t-1}}{i} \alpha^i (1-\alpha)^{Y_{t-1}-i},\ i = 0, 1, 2 \ldots, Y_{t-1} \text{ and } f_2(i) = \frac{\theta^i\, a(i)}{C(\theta)},\ i \in S.$$

Then,

$$\Pr(Y_t = k | Y_{t-1} = l) = f_1 \bullet f_2 = \sum_i f_1(i) f_2(k-i).$$

If $S = \{n, n+1, n+2, \ldots\}$ for fixed $n \in \mathbb{Z}^+$, then

$$0 \leq i \leq l \text{ and } k - i \geq n \Rightarrow 0 \leq i \leq \min(l, k-n),$$

and

$$\Pr(Y_t = k | Y_{t-1} = l) = \sum_{i=0}^{\min(l,k-n)} f_1(i) f_2(k-i).$$

If $S = \{0, 1, 2, \ldots, n\}$ for fixed $n \in \mathbb{Z}^+$, then

$$0 \leq i \leq l \text{ and } 0 \leq k - i \leq n \Rightarrow \max(0, k-n) \leq i \leq \min(l, k),$$

and

$$\Pr(Y_t = k | Y_{t-1} = l) = \sum_{i=\max(0,k-n)}^{\min(l,k)} f_1(i) f_2(k-i).$$

Proof of Proposition 2

Since $\Pr(Y_t = k|Y_{t-1} = l) > 0$, for all k, l, our process is an *irreducible* process, in the sense that every $k \in \mathcal{S}$ can be reached from every $l \in \mathcal{S}$. It also has stationary transition probabilities, in the sense that these transition probabilities do not involve t. Let $P^t(l,k) = \Pr(Y_t = k|Y_0 = l)$. Following Hoel et al. (1972), the existence of a stationary distribution is equivalent to

$$\lim_{t\to\infty} \frac{1}{t} \sum_{m=1}^{t} P^m(x,x) > 0, \qquad x \in \mathcal{S}.$$

From Hoel et al. (1972), if the inequality above is valid for a particular $y \in \mathcal{S}$, it will then, because our process is irreducible, be valid for all $x \in \mathcal{S}$. We will suppose, without loss of generality, that the smallest element of $\mathcal{S}$ is zero, in the sense that $a(0) > 0$ in the power series expansion of $C(\theta)$. Then, we will prove that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{m=1}^{t} P^m(0,0) > 0.$$

This will be proved if we show that $\lim_{m\to\infty} P^m(0,0)$ exists and it is positive. Equivalently, we will show that $\lim_{m\to\infty} \log\left(P^m(0,0)\right)$ exists and is finite.

We begin by showing by induction that

$$P^m(x,0) = \frac{a(0)(1-\alpha^m)^x}{C(\theta)^m} \prod_{i=1}^{m-1} C(\theta(1-\alpha^i)).$$

For $m = 1$, the above expression is reduced to $\Pr(Y_1 = 0|Y_0 = x) = a(0)(1-\alpha)^x/C(\theta)$, which is trivially true. Suppose it is valid for a given m. Then,

$$\begin{aligned}
P^{m+1}(x,0) &= \sum_{z=0}^{\infty} P^1(x,z)P^m(z,0) \\
&= \frac{a(0)}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m-1} C(\theta(1-\alpha^i))\right) \sum_{z=0}^{\infty} \sum_{i=0}^{\min(x,z)} (1-\alpha^m)^z \binom{x}{i} \alpha^i (1-\alpha)^{x-i} \theta^{z-i} a(z-i) \\
&= \frac{a(0)}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m-1} C(\theta(1-\alpha^i))\right) \sum_{i=0}^{x} \sum_{z=i}^{\infty} (1-\alpha^m)^z \binom{x}{i} \alpha^i (1-\alpha)^{x-i} \theta^{z-i} a(z-i) \\
&= \frac{a(0)}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m-1} C(\theta(1-\alpha^i))\right) \sum_{i=0}^{x} (1-\alpha^m)^i \binom{x}{i} \alpha^i (1-\alpha)^{x-i} \sum_{z=i}^{\infty} (1-\alpha^m)^{z-i} \theta^{z-i} a(z-i) \\
&= \frac{a(0)}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m-1} C(\theta(1-\alpha^i))\right) C(\theta(1-\alpha^m)) \sum_{i=0}^{x} (1-\alpha^m)^i \binom{x}{i} \alpha^i (1-\alpha)^{x-i} \\
&= \frac{a(0)}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m} C(\theta(1-\alpha^i))\right) \sum_{i=0}^{x} \left(\frac{1-\alpha^m}{1-\alpha}\right)^i \binom{x}{i} \alpha^i (1-\alpha)^x \\
&= \frac{a(0)(1-\alpha)^x}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m} C(\theta(1-\alpha^i))\right) \sum_{i=0}^{x} \binom{x}{i} \left(1+\alpha+\cdots+\alpha^{m-1}\right)^i \alpha^i \\
&= \frac{a(0)(1-\alpha)^x}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m} C(\theta(1-\alpha^i))\right) \sum_{i=0}^{x} \binom{x}{i} \left(\alpha+\alpha^2+\cdots+\alpha^{m}\right)^i \\
&= \frac{a(0)(1-\alpha)^x}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m} C(\theta(1-\alpha^i))\right) (1+\alpha+\cdots+\alpha^m)^x \\
&= \frac{a(0)(1-\alpha^{m+1})^x}{C(\theta)^{m+1}} \prod_{i=1}^{m} C(\theta(1-\alpha^i)).
\end{aligned}$$

Then, we have

$$\begin{aligned}
P^{m+1}(0,0) &= \sum_{z=0}^{\infty} P^1(0,z)P^m(z,0) \\
&= \frac{a(0)}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m-1} C(\theta(1-\alpha^i)) \right) \sum_{z=0}^{\infty} a(z)\theta^z(1-\alpha^m)^z \\
&= \frac{a(0)}{C(\theta)^{m+1}} \left(\prod_{i=1}^{m-1} C(\theta(1-\alpha^i)) \right) C(\theta(1-\alpha^m)) = \frac{a(0)}{C(\theta)^{m+1}} \prod_{i=1}^{m} C(\theta(1-\alpha^i)) \\
&= \frac{a(0)}{C(\theta)} \prod_{i=1}^{m} \left(\frac{C(\theta(1-\alpha^i))}{C(\theta)} \right).
\end{aligned}$$

Note that for $m = 0$ we have $P^1(0,0) = a(0)/C(\theta)$, which is trivially true. Now, we obtain

$$\log\left(P^{m+1}(0,0)\right) = \log(a(0)) - \log(C(\theta)) - \sum_{i=1}^{m} \left[\log(C(\theta)) - \log\left(C(\theta(1-\alpha^i))\right)\right].$$

Let $G(\theta) = \log(C(\theta))$. From the intermediate value theorem, we can, for each i, obtain a value $\theta_i \in \left((1-\alpha^i)\theta, \theta\right)$ such that $G(\theta) - G(\theta(1-\alpha^i)) = G'(\theta_i)\theta\alpha^i$. Now,

$$\log\left(P^{m+1}(0,0)\right) = \log(a(0)) - G(\theta) - \theta \sum_{i=1}^{m} G'(\theta_i)\alpha^i.$$

Since $G' = C'/C$ is positive, the sum above is a sum of positive terms. It remains, then, to show that the infinite series converges if we let $m \to \infty$. But this is immediate. Observe that $\theta_i \in \left((1-\alpha^i)\theta, \theta\right) \subset [(1-\alpha)\theta, \theta]$. Since G' has a derivative, $G'' = (C''/C) - (G')^2$, G' must be continuous. Let $M(\theta)$ be the maximum value of G' in $[(1-\alpha)\theta, \theta]$. Then,

$$0 \leq \sum_{i=1}^{\infty} G'(\theta_i)\alpha^i \leq M(\theta) \sum_{i=1}^{\infty} \alpha^i = \frac{\alpha M(\theta)}{1-\alpha} < \infty.$$

References

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

Alzaid, A.A. and Al-Osh, M.A. (1988). First-order integer valued autoregressive (INAR(1)) process: distribution and regression properties. *Statistica Neerlandica*, **42**, 53–61.

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

Creel, M.D. and Loomis, J.B. (1990). Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. *American Journal of Agricultural Economics*, **72**, 434–441.

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

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

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

Grogger, J.T. and Carson, R.T. (1991). Models for truncated counts. *Journal of Applied Econometrics*, **6**, 225–238.

Hellström, J. (2001). Unit root testing in integer valued AR(1) models. *Economics Letters*, **70**, 9–14.

Hoel, P.G., Port, S.C. and Stone, C.J. (1972). *Introduction to Stochastic Processes*. Waveland Press.

Jazi, M.A., Jones, G. and Lai, C.D. (2012a). Integer Valued AR(1) with Geometric Innovations. *Journal of the Iranian Statistical Society*, **11**, 173–190.

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

Johnson, N.L., Kemp, A.K. and Kotz, S. (2005). *Univariate Discrete Distributions 3º ed.*. Wiley.

Jung, R.C. and Tremayne, A.R. (2003). Testing for serial dependence in time series models of counts. *Journal of Time Series Analysis*, **24**, 65–84.

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

Klimko, L. and Nelson, P. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistic*, **6**, 629–642.

Nastić, A.S., Ristić, M.M. and Bakouch, H.S (2012). A combined geometric INAR(p) model based on negative binomial thinning. *Mathematical and Computer Modelling*, **55**, 1665–1672.

Noack, A. (1950). A Class of random variables with discrete distributions. *Annals of Mathematical Statistics* **21**, 127–132.

Rajarshi, M.B. (2012). *Statistical Inference for Discrete Time Stochastic Processes*. Springer Verlag.

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.

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

Sprott, D.A. (1983). Estimating the parameters of a convolution by maximum likelihood. *Journal of the American Statistical Association*, **78**, 457–460.

Steutel, F.W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, **7**, 893–899.

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

Weiß, C.H. (2009). A new class of autoregressive models for time series of binomial counts. *Communications in Statistics - Theory and Methods*, **38**, 447–460.

Weiß, C.H. (2011). Simultaneous confidence regions for the parameters of a Poisson INAR(1) model. *Statistical Methodology*, **8**, 517–527.

Zhang, H, Wang, D. and Zhu, F. (2010). Inference for INAR(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference*, **140**, 667–683.

Zhu, F. (2012). Zero inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference*, **142**, 826–839.

CHAPTER 3

Improved estimation for Poisson INAR(1) models

Resumo

Nesse capítulo, consideramos o modelo Poisson INAR(1) proposto por McKenzie (1985) e Al-Osh e Alzaid (1987), e obtemos o viés de segunda ordem do estimador diferença de quadrados e utilizamos esse viés para definir um estimador com viés reduzido. O comportamento do estimador de mínimos quadrados condicional corrigido também é estudado. Além disso, encontramos a distribuição assintótica dos estimadores propostos e comparamos os desempenhos dos estimadores propostos por meio de um extensivo estudo de simulação. Um uso prático deste modelo e das fórmulas obtidas para correção de viés é apresentado.

Palavras-chave: Correção de viés. Estimador de mínimos quadrados condicional. Estimador de máxima verossimilhança condicional. Estimador diferença de quadrados. Processo Poisson INAR(1).

Abstract

We consider the first-order Poisson autoregressive model proposed by McKenzie (1985) and Al-Osh and Alzaid (1987), which is suitable to situations where the time series data is non-negative and integer valued. We derive the second order bias of the squared difference estimator for one of the parameters and use that bias to define a bias-adjusted estimator. The behaviour of a modified conditional least squares estimator is also studied. Further, we access the asymptotic properties of the estimators here discussed. We present numerical evidence, based upon Monte Carlo simulation studies, showing that the proposed bias-adjusted estimator outperforms the other estimators in small samples. We also present an application to a real data set.

Keywords: Bias correction. Conditional least squares estimator. Conditional maximum likelihood estimator. Poisson INAR(1) process. Squared difference estimator.

3.1 Introduction

McKenzie (1985) proposed the first-order non-negative integer-valued autoregressive (INAR(1)) process with Poisson innovations to model discrete-valued time series, and independently little later once more by Al-Osh and Alzaid (1987). The main motivation for this model is the need to model series with correlated observations expressed as (small) counts. Examples of these series are the daily counts of epileptic seizure in one patient (Franke and Seligmann, 1993), the number of generics for different medical (Hellström, 2001), the number of guest nights in hotels (Brännäs et al., 2002), the number of different IP addresses (Weiß, 2007).

Many new results on Poisson INAR(1) models have been obtained in recent years. For example, Park and Oh (1997) studied the asymptotic properties of Yule-Walker (YW) estimators of the parameters. Hellström (2001) focused on the testing of a unit root. Freeland and McCabe (2005) studied the asymptotic properties of the conditional least squares (CLS) estimators. Weiß (2007) investigated the distribution of the process of jumps from a Poisson INAR(1) process. Weiß (2011) proposed several asymptotic simultaneous confidence regions for the parameters.

The most widely used estimators in the literature for the parameters of Poisson INAR(1) processes are YW, CLS and conditional maximum likelihood (CML) estimators. It is well known that all those estimators are biased in finite samples, yet little has been done in practice to produce estimators with smaller biases. The derivation of alternative estimators with smaller biases is particularly important when the number of observations are relatively small, a situation for which biases of the traditionally used estimators can become large.

In this context, the main goal of this chapter is to introduce a new estimator with a smaller bias than the traditionally used ones. With this aim, we have derived a closed-form expression for the second order bias of the squared difference (SD) estimator of one of the parameters and used this expression to construct a bias-corrected estimator. We have also considered an alternative estimator based on the results in Jung et al. (2005), whose asymptotic distribution is also obtained.

The chapter unfolds as follows. In Section 3.2, the Poisson INAR(1) process is introduced, some of its basic properties are outlined and CML is described. In Section 3.3, we derive the second order biases of the SD estimator, present a possible alternative based on CLS estimator and study asymptotic properties of the estimators. Numerical results from Monte Carlo simulation experiments are presented and discussed in Section 3.4. In Section 3.5, we consider an empirical example. Finally, Section 3.6 concludes the chapter.

3.2 The Poisson INAR(1) process

Let $\mathcal{D}(\mathbb{Z}^+)$ denote the class of all probability distributions having as support the set $\mathbb{Z}^+$ of nonnegative integers. We consider the binomial thinning operator $\circ : [0,1] \times \mathcal{D}(\mathbb{Z}^+) \to \mathcal{D}(\mathbb{Z}^+)$. For each real number $\alpha \in [0,1]$ and for each probability distribution π over $\mathbb{Z}^+$, the distribution $\circ(\alpha, \pi)$, also denoted as $\alpha \circ \pi$, is defined as follows. Let $\{U_n; n \geq 1\}$ be a sequence of independent and identically distributed (i.i.d.) Bernoulli random variables with probability $P(U_n = 1) = \alpha = 1 - P(U_n = 0)$. Let X be a nonnegative random variable, independent of that sequence, with distribution π. The distribution $\alpha \circ \pi$ is the distribution of the random sum $U_1 + \cdots + U_X$. Also, if X is a nonnegative random variable with distribution π, we can informally use $\alpha \circ X$ to denote $\alpha \circ \pi$. Clearly, the name "thinning" comes from the fact that $U_1 + \cdots + U_X \leq X$.

It is easy to obtain some elementary properties of the binomial thinning operator. If π is any probability distribution over $\mathbb{Z}^+$, then, for example, $0 \circ \pi = \delta_0$, the probability distribution concentrated at 0. Also, $1 \circ \pi = \pi$. In general, if ψ is the probability generating function for π, then, the probability generating function $\widetilde{\psi}$ for $\alpha \circ \pi$ will be given by $\widetilde{\psi}(t) = \psi(\alpha t + 1 - \alpha)$. Furthermore, using probability generating functions, we can easily conclude that the distribution $\alpha \circ (\beta \circ \pi)$, where α and β are in $[0,1]$, is the distribution $(\alpha\,\beta) \circ \pi$. Concerning expected values, it is also trivial to verify that $\mathrm{E}(\alpha \circ \pi) = \alpha \mathrm{E}(\pi)$ and $\mathrm{Var}[\alpha \circ \pi] = \alpha^2 \mathrm{Var}(\pi) + \alpha(1-\alpha)\mathrm{E}(\pi)$. Furthermore, if π_1 and π_2 are probability distributions in $\mathcal{D}(\mathbb{Z}^+)$ with finite variances, it comes easily that for any $\alpha, \beta \in [0,1]$ we have $\mathrm{Cov}(\alpha \circ \pi_1, \beta \circ \pi_2) = \alpha\,\beta\,\mathrm{Cov}(\pi_1, \pi_2)$. For an account of the properties of the binomial thinning operator, see Silva and Oliveira (2004).

McKenzie (1985) and Al-Osh and Alzaid (1987) introduced the INAR(1) process, the integer-valued autoregressive process of first order, which is recursively defined as

$$Y_t \overset{d}{=} \alpha \circ Y_{t-1} + \epsilon_t, t \in \mathbb{Z},$$

where ϵ_t is a sequence of i.i.d. random variables with common distribution in $\mathcal{D}(\mathbb{Z}^+)$. The symbol $\overset{d}{=}$ means that the distribution of Y_t is the distribution of $\alpha \circ Y_{t-1} + \epsilon_t$. It is important to observe that, if $0 < \alpha < 1$, then, Y_t is not uniquely defined from Y_{t-1} and ϵ_t, a situation that differs from that of the classical AR(1) model $X_t = \alpha\, X_{t-1} + u_t$.

The i.i.d. property of the sequence $\{\epsilon_t\}_{t\in\mathbb{Z}}$ implies that Y_t is a time-homogeneous Markov chain. Also, if $\alpha < 1$ and ϵ_t is Poisson distributed with $\mathrm{E}(\epsilon_t) = \lambda$, then the Poisson distribution with parameter $\lambda/(1-\alpha)$ is a stationary distribution for the process. It is not difficult to verify that $\mathrm{E}\left(Y_{t+k}|Y_t\right) = \alpha^k Y_t + \lambda(1-\alpha^k)/(1-\alpha)$ for $k > 0$. From this fact, it follows immediately that the correlation function $\rho(k) = \mathrm{Corr}(Y_t, Y_{t-k})$, for $k > 0$, is then given by $\rho(k) = \alpha^k$, a correlation function that decays exponentially with the lag, exactly as happens in the classical AR(1) model $X_t = \alpha\, X_{t-1} + u_t$, where u_t is a white noise with finite variance. Hereafter, the sequence of i.i.d. random variables ϵ_t is Poisson distributed with $\mathrm{E}(\epsilon_t) = \lambda$ and $\{Y_t\}_{t\in\mathbb{Z}}$ is a stationary Markov chain with Poisson marginal distribution with parameter $\lambda/(1-\alpha)$.

The expected value, variance, skewness and kurtosis are, respectively,

$$\mathrm{E}(Y_t) = \frac{\lambda}{1-\alpha} = \mu,\ \mathrm{Var}(Y_t) = \mu,\ \gamma_1 = 1/\sqrt{\mu},\ \gamma_2 = 3 + \mu^{-1}.$$

Note, therefore, that the Poisson INAR(1) model has the properties

$$\mathrm{E}(Y_t) = \mathrm{Var}(Y_t)\ \text{ and }\ \gamma_2 - \gamma_1^2 - 3 = 0.$$

The kth raw moment of $\{Y_t\}_{t\in\mathbb{Z}}$ is recursively given by

$$\mathrm{E}(Y_t^k) = \mu_k = \mu\left(\mu_{k-1} + \frac{d\mu_{k-1}}{d\mu}\right), k \geq 1.$$

In particular,

$$\mathrm{E}(Y_t^2) = \mu_2 = \mu + \mu^2,\, \mathrm{E}(Y_t^3) = \mu_3 = \mu + 3\mu^2 + \mu^3 \text{ and } \mathrm{E}(Y_t^4) = \mu_4 = \mu^4 + 6\mu^3 + 7\mu^2 + \mu.$$

Strict stationarity of Poisson INAR(1) implies that the kth-order joint moment of the random variables $Y_t, Y_{t+s_1}, \ldots, Y_{t+s_{k-1}}$ exists and has the form

$$\mathrm{E}(Y_t Y_{t+s_1} \ldots Y_{t+s_{k-1}}) = \mu(s_1, \ldots, s_{k-1}).$$

Formulae for second-order, third-order and fourth-order joint moments of the Poisson INAR(1) process are given below. This result can be found in Silva and Oliveira (2004) and Weiß (2012).

Lemma 1. *The second-order, third-order and fourth-order joint moments of the Poisson INAR(1) process are given by*

1. $\mu(s) = \mu\,\alpha^s + \mu^2,\ s \geq 0,$
2. $\mu(s,u) = \mu^3 + \mu(1+\mu)\alpha^u + \mu^2\alpha^s + \mu^2\alpha^{u-s},\ u \geq s \geq 0,$
3. $\mu(s,u,l) = \mu^2(\alpha^{l-u+s} + 2\alpha^{l+u-s}) + \mu(1+\mu)^2\alpha^l + \mu^2(1+\mu)(\alpha^{l-s} + \alpha^u) + \mu^3(\alpha^{l-u} + \alpha^{u-s} + \alpha^s) + \mu^4,\ l \geq u \geq s \geq 0.$

The transition probabilities of this process are given by

$$\Pr(Y_t = k | Y_{t-1} = l) = \mathrm{e}^{-\lambda} \sum_{i=0}^{\min(k,l)} \frac{\lambda^{k-i}}{(k-i)!} \binom{l}{i} \alpha^i (1-\alpha)^{l-i}, \tag{3.1}$$

where $\binom{\cdot}{\cdot}$ is the standard combinatorial symbol. For additional properties of a stationary Poisson INAR(1) process, see Silva and Oliveira (2004).

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. Among the most used methods for parameter estimation is conditional maximum likelihood (CML). The CML estimator of $\boldsymbol{\eta} = (\alpha,\lambda)^\top$ is the value $\widehat{\boldsymbol{\eta}} = (\widehat{\alpha},\widehat{\lambda})^\top$ that maximizes the conditional log-likelihood function $\ell(\alpha,\lambda)$. Suppose that y_1 is fixed. The conditional log-likelihood function for the Poisson INAR(1) model is defined as

$$\ell(\alpha,\lambda) = \log\left[\prod_{t=2}^{n} \Pr(y_t|y_{t-1})\right] = \sum_{t=2}^{n} \log[\Pr(y_t|y_{t-1})],$$

where $\Pr(y_t|y_{t-1})$ is given in (3.1). CML estimators will be obtained, in general, using numerical methods, since equating the first-order log-likelihood derivatives to zero leads us to a complicated system of nonlinear equations.

Proposition 1. *(Freeland and McCabe, 2004). The CML estimators are asymptotically normal, i.e.*

$$\sqrt{T}(\widehat{\alpha} - \alpha, \widehat{\lambda} - \lambda) \xrightarrow{d} \mathrm{N}(0, \mathbf{K}^{-1}),$$

where $\mathbf{K}$ *is the Fisher information matrix.*

CML estimators, as commented above, may not have closed-form expressions and the Fisher's information matrix is not available. The alternative estimators discussed in the literature (YW and CLS), in contrast, can be very easily computed and such simple procedures are often attractive. However, these alternative estimators perform worse in terms of bias and mean square error than CML (Al-Osh and Alzaid, 1987). In this context, a fourth method of estimation, the SD estimator, was proposed. From a sample $Y_1,\ldots,Y_T$ of consecutive observations of a Poisson INAR(1) process with parameters α and λ, we have

$$\mathrm{E}\left[\sum_{t=2}^{T} (Y_t - Y_{t-1})^2\right] = 2(T-1)\lambda,$$

and, therefore, an unbiased estimator for λ can be obtained as

$$\widetilde{\lambda} = \frac{1}{2(T-1)} \sum_{t=2}^{T} (Y_t - Y_{t-1})^2, \tag{3.2}$$

the SD estimator. Weiß (2012) introduces this SD estimator of λ for a Poisson INAR(1) process and derive its exact as well as asymptotic stochastic properties. The SD estimator of α is, then, based on the first moment of $\{Y_t\}_{t\in\mathbb{Z}}$, which is $\mathrm{E}(Y_t) = \lambda/(1-\alpha)$. It is be given by

$$\widetilde{\alpha} = 1 - \frac{\widetilde{\lambda}}{\overline{Y}},$$

where $\widetilde{\lambda}$ is the SD estimator defined in (3.2) and $\overline{Y} = T^{-1}\sum_{t=1}^{T} Y_t$. Note that $\widetilde{\alpha}$ is obtained from two unbiased estimators, since $\mathrm{E}(\widetilde{\lambda}) = \lambda$ and $\mathrm{E}(\overline{Y}) = \mu$.

3.3 Asymptotic properties of estimators

This section deals with bias reduction of some of the aforementioned estimators of α. Although the SD estimator for λ is unbiased, the same is not true for the SD estimator of α. For a large sample size T, the bias does not typically represent a problem, since its order is, in general, $O(T^{-1})$, while the standard deviation of the same estimator is, in general, of order $O(T^{-1/2})$. However, when the sample size is not large, the bias can get considerably large and techniques for reducing bias will become relevant.

We start with the proposition below, whose proof can be found in the Appendix.

Proposition 2. *The asymptotic joint distribution of the SD estimators* $\sqrt{T}(\widetilde{\alpha} - \alpha, \widetilde{\lambda} - \lambda)$ *is* $\mathrm{N}(0, \mathbf{\Sigma})$, *where*

$$\mathbf{\Sigma} = \begin{pmatrix} \frac{\alpha(1-\alpha)^2}{\lambda} + (1-\alpha)^2\frac{3+\alpha}{1+\alpha} & -\lambda(1-\alpha)\frac{3+\alpha}{1+\alpha} \\ & \\ -\lambda(1-\alpha)\frac{3+\alpha}{1+\alpha} & \lambda\left(1+\lambda\frac{3+\alpha}{1+\alpha}\right) \end{pmatrix}.$$

We now consider a second order Taylor expansion of $\widetilde{\alpha} = 1 - \frac{\widetilde{\lambda}}{\overline{Y}}$ around $\boldsymbol{\eta} = (\lambda, \mu)^{\mathsf{T}}$. From the above proposition, we know that $\overline{Y} - \mu = O_P(T^{-1/2})$ and $\widetilde{\lambda} - \lambda = O_P(T^{-1/2})$. Therefore,

$$\frac{\widetilde{\lambda}}{\overline{Y}} = \frac{\lambda}{\mu} + \mu^{-1}(\overline{Y} - \mu) - \frac{\lambda}{\mu^2}(\widetilde{\lambda} - \lambda) - \mu^{-2}(\widetilde{\lambda} - \lambda)(\overline{Y} - \mu) + \frac{\lambda}{\mu^3}(\overline{Y} - \mu)^2 + O_P(T^{-3/2}),$$

thus implying that the bias of $\widetilde{\alpha}$, $B(\widetilde{\alpha}) = \mathrm{E}(\widetilde{\alpha} - \alpha)$, is

$$B(\widetilde{\alpha}) = \mu^{-2}\left[\mathrm{Cov}(\widetilde{\lambda}, \overline{Y}) - \frac{\lambda}{\mu}\mathrm{Var}(\overline{Y})\right] + o(T^{-1}).$$

From the definition of $\overline{Y}$, we have

$$\begin{aligned} \mathrm{Var}(\overline{Y}) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}(Y_t, Y_s) = \frac{1}{T^2}\left[T\mu + 2\sum_{s=1}^{T}\sum_{t>s}^{T}\mathrm{Cov}(Y_t, Y_s)\right] \\ &= \frac{1}{T^2}\left[T\mu + 2\,\mu\sum_{s=1}^{T}\sum_{t>s}^{T}\rho(t-s)\right] = \frac{1}{T^2}\left[T\mu + 2\,\mu\sum_{s=1}^{T}\sum_{t>s}^{T}\alpha^{t-s}\right] \\ &= \frac{\mu}{T^2}\left[T + 2\sum_{s=1}^{T}\sum_{i=1}^{T-s}\alpha^i\right] = \frac{\mu}{T^2}\left[T + 2\left(\frac{T\alpha}{1-\alpha} - \frac{\alpha(1-\alpha^T)}{(1-\alpha)^2}\right)\right] \\ &= \frac{\mu}{T} + \frac{2\,\alpha\,\mu}{T(1-\alpha)} - \frac{2\,\alpha\,\mu(1-\alpha^T)}{T^2(1-\alpha)^2} = \frac{\mu(1+\alpha)}{T(1-\alpha)} - \frac{2\,\alpha\,\mu(1-\alpha^T)}{T^2(1-\alpha)^2}. \end{aligned}$$

Thus,

$$\frac{\lambda}{\mu^3}\mathrm{Var}(\overline{Y}) = \frac{1}{T}\frac{1+\alpha}{\mu} + O(T^{-2}).$$

On the other hand, $\mathrm{Cov}(\widetilde{\lambda}, \overline{Y}) = \mathrm{E}(\widetilde{\lambda}\,\bar{y}) - \mathrm{E}(\widetilde{\lambda})\mathrm{E}(\overline{Y}) = \mathrm{E}(\widetilde{\lambda}\,\overline{Y}) - \lambda\,\mu$, where

$$\mathrm{E}(\widetilde{\lambda}\,\overline{Y}) = \mathrm{E}\left[\frac{1}{2(T-1)}\sum_{s=2}^{T}(Y_s - Y_{s-1})^2\frac{1}{T}\sum_{t=1}^{T}Y_t\right] = \frac{1}{2\,T(T-1)}\sum_{s=2}^{T}\sum_{t=1}^{T}\mathrm{E}\left[Y_t(Y_s - Y_{s-1})^2\right].$$

Further,

$$\mathrm{E}\left[Y_t(Y_s - Y_{s-1})^2\right] = \begin{cases} \mu(0,t-s) + \mu(0,t-s+1) - 2\mu(1,t-s+1), & \text{if } t \geq s \\ \mu(s-t,s-t) + \mu(s-t-1,s-t-1) - 2\mu(s-t-1,s-t), & \text{if } t < s, \end{cases}$$

and then, using Lemma 1,

$$\mathrm{E}\left[Y_t(Y_s - Y_{s-1})^2\right] = \begin{cases} 2\,\lambda\,\mu + \lambda\,\alpha^{t-s}, & \text{if } t \geq s \\ 2\,\lambda\,\mu + \lambda\,\alpha^{s-t-1}, & \text{if } t < s. \end{cases}$$

Applying the geometric sum, we have

$$\begin{aligned}
\mathrm{E}(\widetilde{\lambda}\,\overline{Y}) &= \frac{1}{2\,T(T-1)} \sum_{s=2}^{T}\sum_{t=1}^{T} \mathrm{E}\left[Y_t(Y_s - Y_{s-1})^2\right] \\
&= \frac{1}{2\,T(T-1)} \sum_{s=2}^{T}\sum_{t=s}^{T} \left[2\,\lambda\,\mu + \lambda\,\alpha^{t-s}\right] + \frac{1}{2\,T(T-1)} \sum_{s=2}^{T}\sum_{t<s}^{T} \left[(\lambda/\alpha)\alpha^{s-t} + 2\,\lambda\,\mu\right] \\
&= \frac{\lambda\,\mu}{2} + \frac{\lambda}{2\,T(T-1)} \sum_{s=2}^{T}\sum_{j=0}^{T-s} \alpha^j + \frac{\lambda}{2\,\alpha\,T(T-1)} \sum_{s=2}^{T}\sum_{j=1}^{s-1} \alpha^j + \frac{\lambda\,\mu}{2} \\
&= \lambda\,\mu + \frac{\lambda}{2\,T(T-1)} \left[\frac{T-1}{1-\alpha} - \frac{\alpha(1-\alpha^{T-1})}{(1-\alpha)^2}\right] + \frac{\lambda}{2\,\alpha\,T(T-1)} \left[\frac{\alpha(T-1)}{1-\alpha} - \frac{\alpha^2(1-\alpha^{T-1})}{(1-\alpha)^2}\right] \\
&= \lambda\,\mu + \frac{\mu}{T} - \frac{\alpha\,\mu(1-\alpha^{T-1})}{T(T-1)(1-\alpha)}.
\end{aligned}$$

Thus,

$$\mu^{-2}\mathrm{Cov}(\widetilde{\lambda}, \overline{Y}) = \frac{1}{T\,\mu} + O(T^{-2}).$$

So, we obtain

$$B(\widetilde{\alpha}) = -\frac{\alpha}{T\,\mu} + o(T^{-1}). \tag{3.3}$$

Using (3.3), we consider a bias-reduced SD estimator $\overline{\alpha}$ of α as

$$\overline{\alpha} = \widetilde{\alpha} + \frac{\widetilde{\alpha}}{T\,\overline{Y}},$$

i.e., the unknown parameters in (3.3) are replaced by their corresponding estimators.

It remains to show that the bias of this new estimator $\overline{\alpha}$ is indeed $o(T^{-1})$. Note that

$$\frac{\widetilde{\alpha}}{\overline{Y}} = \frac{\alpha}{\mu} + \mu^{-1}(\overline{Y} - \mu) - \frac{\alpha}{\mu^2}(\widetilde{\alpha} - \alpha) - \mu^{-2}(\widetilde{\alpha} - \alpha)(\overline{Y} - \mu) + \frac{\alpha}{\mu^3}(\overline{Y} - \mu)^2 + O_P(T^{-3/2}).$$

Therefore,

$$\begin{aligned}
\mathrm{E}\left(\frac{\widetilde{\alpha}}{\overline{Y}}\right) &= \frac{\alpha}{\mu} - \frac{\alpha}{\mu^2}B(\widetilde{\alpha}) - \mu^{-2}\mathrm{E}\left[(\widetilde{\alpha} - \alpha)(\overline{Y} - \mu)\right] + \frac{\alpha}{\mu^3}\mathrm{Var}(\overline{Y}) + o(T^{-1}) \\
&= \frac{\alpha}{\mu} - \mu^{-2}\mathrm{E}\left(\widetilde{\alpha}(\overline{Y} - \mu)\right) + O(T^{-1}) \\
&= \frac{\alpha}{\mu} - \mu^{-2}\mathrm{E}\left(\widetilde{\alpha}\overline{Y}\right) + \mu^{-1}\mathrm{E}\left(\widetilde{\alpha}\right) + O(T^{-1}) \\
&= \frac{\alpha}{\mu} - \mu^{-2}\mathrm{E}\left(\overline{Y} - \widetilde{\lambda}\right) + \frac{\alpha}{\mu} + O(T^{-1}) \\
&= \frac{\alpha}{\mu} + \frac{\lambda - \mu}{\mu^2} + \frac{\alpha}{\mu} + O(T^{-1}) \\
&= \frac{\alpha}{\mu} + \frac{\mu(1-\alpha) - \mu}{\mu^2} + \frac{\alpha}{\mu} + O(T^{-1}) = \frac{\alpha}{\mu} + O(T^{-1}).
\end{aligned}$$

It then follows that

$$\begin{aligned} \mathrm{E}(\overline{\alpha}) &= \mathrm{E}\left(\widetilde{\alpha} + \frac{\widetilde{\alpha}}{T\overline{Y}}\right) \\ &= \mathrm{E}\left(\widetilde{\alpha}\right) + \frac{1}{T}\mathrm{E}\left(\frac{\widetilde{\alpha}}{\overline{Y}}\right) \\ &= \alpha - \frac{\alpha}{T\mu} + o(T^{-1}) + \frac{1}{T}\left[\frac{\alpha}{\mu} + O(T^{-1})\right] \\ &= \alpha + o(T^{-1}) + O(T^{-2}) = \alpha + o(T^{-1}), \end{aligned}$$

as we wanted. Therefore, $\overline{\alpha}$ is indeed a bias-adjusted estimator up to order $O(T^{-1})$. It is easy to see that $\overline{\alpha}$ has the same asymptotic normal distribution as the original estimator $\widetilde{\alpha}$, as it usually happens with bias-corrected estimators. The simple argument involves the fact that the original estimator is $O_P(T^{-1/2})$ while its bias is $O(T^{-1})$.

We consider yet another estimator. Kendall (1954) as well as Mariott and Pope (1954) derived the approximate bias of the conditional least squares (CLS) estimator of the parameter α in the AR(1) model with drift $X_t = \lambda + \alpha X_{t-1} + u_t$, where u_t is an i.i.d. random sequence with zero mean. They obtained

$$B(\breve{\alpha}) = -\frac{1}{T}(1 + 3\alpha) + O(T^{-2}), \tag{3.4}$$

where $\breve{\alpha}$ is the CLS estimator

$$\breve{\alpha} = \frac{\sum_{t=2}^{T} Y_t Y_{t-1} - \frac{1}{T-1}\sum_{t=2}^{T} Y_t \sum_{t=2}^{T} Y_{t-1}}{\sum_{t=2}^{T} Y_{t-1}^2 - \frac{1}{T-1}\left(\sum_{t=2}^{T} Y_{t-1}\right)^2}.$$

Orcutt and Winokur (1969) used (3.4) to propose the bias-reduced estimator

$$\check{\alpha} = \frac{1}{T-3}(T\breve{\alpha} + 1), \tag{3.5}$$

whose bias is $O(T^{-2})$. The corresponding estimator $\check{\lambda}$ arises naturally from using $\check{\alpha}$ rather than $\breve{\alpha}$ in its expression, i.e.

$$\check{\lambda} = \frac{1}{T-1}\left(\sum_{t=2}^{T} Y_t - \check{\alpha}\sum_{t=2}^{T} Y_{t-1}\right).$$

Since CLS for the Gaussian AR(1) model has exactly the same expression as for the Poisson INAR(1) model (because the conditional mean is the same) and since the autocovariance structure is also identical, it may be of interest to see if bias correction methods used in the Gaussian context also improve estimation performance in the Poisson INAR(1) case. We, therefore, consider the bias corrected estimator in (3.5).

The proof of the proposition below can be found in the Appendix.

Proposition 3. *For the estimator* $\check{\boldsymbol{\eta}} = (\check{\alpha}, \check{\lambda})^{\mathsf{T}}$, *in the Poisson INAR(1) process, we have*

$$\sqrt{T}(\check{\alpha} - \alpha) \xrightarrow{d} N\left(0, \frac{\alpha(1-\alpha)^2}{\lambda} + (1-\alpha)(1+\alpha)\right)$$

and

$$\sqrt{T}(\check{\lambda} - \lambda) \xrightarrow{d} N\left(0, \lambda\left(1 + \frac{1+\alpha}{1-\alpha}\lambda\right)\right).$$

3.4 Numerical evaluation

The aim of the simulation study presented in this section is to examine the small sample properties of the different estimators previously described and compare their behaviours. We have evaluated, through Monte Carlo simulations, the finite sample performances of the CML $\widehat{\boldsymbol{\eta}} = (\widehat{\alpha}, \widehat{\lambda})^\top$, SD $\widetilde{\boldsymbol{\eta}} = (\widetilde{\alpha}, \widetilde{\lambda})^\top$, modified CLS $\breve{\boldsymbol{\eta}} = (\breve{\alpha}, \breve{\lambda})^\top$ and corrected SD $\overline{\alpha}$ estimators for the parameters α and λ of the Poisson INAR(1) process.

The data set $Y_1, \ldots, Y_T$ is were generated from a Poisson INAR(1) process with $\{\epsilon_t\}$ being an i.i.d. Poisson sequence with mean λ. The sample sizes considered were $T = 15, 30, 45, 60$ and the values of the α parameter in the simulation study were $\alpha = 0.35, 0.50, 0.65, 0.80$. We set values of λ at 1.0, 3.0 and 5.0. For each different situation the bias and mean squared error (MSE) of the estimators are numerically estimated.

Although it was always possible to compute the estimates according to the formulae given in Sections 3.2-3.4, the obtained estimates did not always satisfy the restrictions on (α, λ) induced by a Poisson INAR(1) process. The class of conditional likelihood-based estimators is the only group of estimators explicitly able to prevent inadmissible estimates using a suitable parameter transformation. Here, we consider $\beta = \ln[\alpha/(1-\alpha)]$ and $\delta = \ln(\lambda)$. The resulting estimates for α and λ were then transformed back via: $\alpha = \exp(\beta)/[1 + \exp(\beta)]$ and $\lambda = \exp(\delta)$. For other estimators, if $(\widetilde{\alpha}, \widetilde{\lambda}), (\overline{\alpha}, \widetilde{\lambda}), (\breve{\alpha}, \breve{\lambda}) \notin (0,1) \times (0, \infty)$ repeat the replica.

The Monte Carlo simulation experiments were performed using the R environment; see `http://www.r-project.org`. The number of Monte Carlo replications for each situation is $R = 5000$. The CML estimators of α and λ were obtained by maximizing the conditional log-likelihood function using the quasi-Newton nonlinear optimization method BFGS with numerical derivatives. As starting values for the algorithm, we suggest the estimates obtained by CLS method.

Tables 3.1, 3.2 and 3.3 give, for $\lambda = 1$, $\lambda = 3$ and $\lambda = 5$, respectively, the estimated biases for the four estimators of α and the three estimators of λ considered in our study. We observe that, for all situations, CML estimators of α and λ have the worst performance in terms of bias. Also, the best estimator of α, in terms of bias, is, for all situations, our proposed estimator, the corrected SD estimator $\overline{\alpha}$, while $\widetilde{\alpha}$ and $\breve{\alpha}$ seem to compete.

Although we know that $\widetilde{\lambda}$ is an unbiased estimator of λ, we decide also to estimate its bias in our simulation study, in order to compare with the biases of the other two estimators. The comparison between the estimators of λ shows that $\breve{\lambda}$ has indeed a much better performance than CML and its bias can become very small in some situations. There are situations, however, where its bias will not be so small.

In order to summarize the results obtained in our simulations, we introduce a quantity that we call the integrated relative absolute bias (IRAB). Let η be an unknown parameter to be estimated (η may be α or λ) and suppose that η^* is any estimator of η. Suppose that we estimate, by simulation, the bias of η^* as estimator of η for N different situations (in our particular case, $N = 12$, since we considered four different values of α and three different values of λ). Let η_j be the value of η for situation j (observe that some of the η_j will be equal) and let b_j be the estimated bias of η^* for this situation. We define the IRAB as

$$IRAB = \sum_{j=1}^{N} \frac{|b_j|}{\eta_j}.$$

Table 3.4 presents the IRAB for the seven estimators and the four sample sizes considered. This table confirms the fact that the CML estimators of α and λ present the worst performances in terms of bias and that our corrected SD estimator is uniformly better than the other estimators of α in terms of

Table 3.1: Biases ($\lambda = 1.0$).

T	α	Estimator of α				Estimator of λ		
		$\widehat{\alpha}$	$\widetilde{\alpha}$	$\breve{\alpha}$	$\overline{\alpha}$	$\widehat{\lambda}$	$\widetilde{\lambda}$	$\breve{\lambda}$
15	0.35	−0.0513	−0.0118	−0.0060	0.0036	0.0622	−0.0030	−0.0113
	0.50	−0.0633	−0.0137	−0.0085	0.0034	0.1072	0.0078	−0.0022
	0.65	−0.0469	−0.0107	−0.0117	0.0054	0.0905	−0.0067	0.0038
	0.80	−0.0233	−0.0078	−0.0267	0.0036	0.0743	0.0044	0.1107
30	0.35	−0.0317	−0.0056	0.0016	0.0019	0.0386	−0.0020	−0.0135
	0.50	−0.0344	−0.0116	−0.0051	−0.0032	0.0490	0.0070	−0.0096
	0.65	−0.0251	−0.0066	−0.0077	0.0012	0.0483	−0.0016	0.0021
	0.80	−0.0120	−0.0032	−0.0159	0.0024	0.0313	−0.0091	0.0600
45	0.35	−0.0252	−0.0054	−0.0034	−0.0004	0.0306	0.0005	−0.0029
	0.50	−0.0230	−0.0053	−0.0062	0.0003	0.0299	−0.0044	−0.0024
	0.65	−0.0150	−0.0033	−0.0064	0.0019	0.0256	−0.0072	0.0025
	0.80	−0.0080	−0.0022	−0.0053	0.0015	0.0219	−0.0054	0.0098
60	0.35	−0.0184	−0.0035	−0.0029	0.0003	0.0257	0.0028	0.0017
	0.50	−0.0177	−0.0040	−0.0064	0.0002	0.0245	−0.0019	0.0019
	0.65	−0.0116	−0.0032	−0.0032	0.0007	0.0187	−0.0046	−0.0052
	0.80	−0.0068	−0.0021	−0.0034	0.0006	0.0193	−0.0023	0.0053

$\widehat{\eta}$: CML; $\widetilde{\eta}$: SD; $\breve{\eta}$: modified CLS; $\overline{\eta}$: corrected SD.

Table 3.2: Biases ($\lambda = 3.0$).

T	α	Estimator of α				Estimator of λ		
		$\widehat{\alpha}$	$\widetilde{\alpha}$	$\breve{\alpha}$	$\overline{\alpha}$	$\widehat{\lambda}$	$\widetilde{\lambda}$	$\breve{\lambda}$
15	0.35	−0.0549	−0.0172	−0.0151	−0.0123	0.2294	0.0596	0.0422
	0.50	−0.0573	−0.0086	−0.0121	−0.0030	0.3060	0.0164	0.0365
	0.65	−0.0372	−0.0039	−0.0160	0.0013	0.2608	−0.0119	0.0907
	0.80	−0.0126	−0.0020	−0.0280	0.0017	0.1380	−0.0179	0.3879
30	0.35	−0.0334	−0.0046	−0.0067	−0.0021	0.1508	0.0165	0.0267
	0.50	−0.0273	−0.0038	−0.0032	−0.0010	0.1386	0.0021	−0.0042
	0.65	−0.0174	−0.0020	−0.0047	0.0005	0.1221	−0.0052	0.0140
	0.80	−0.0091	−0.0024	−0.0110	−0.0006	0.1144	0.0158	0.1480
45	0.35	−0.0234	−0.0024	−0.0030	−0.0008	0.1001	0.0018	0.0063
	0.50	−0.0180	−0.0030	−0.0015	−0.0011	0.0872	−0.0008	−0.0119
	0.65	−0.0112	−0.0016	0.0001	0.0001	0.0844	0.0033	−0.0126
	0.80	−0.0055	−0.0008	−0.0056	0.0004	0.0668	−0.0056	0.0718
60	0.35	−0.0170	−0.0017	−0.0023	−0.0004	0.0729	0.0026	0.0053
	0.50	−0.0119	−0.0016	0.0002	−0.0002	0.0648	0.0041	−0.0084
	0.65	−0.0087	−0.0014	−0.0004	0.0001	0.0653	0.0028	−0.0001
	0.80	−0.0047	0.0009	0.0051	−0.0001	0.0539	−0.0038	0.0614

$\widehat{\eta}$: CML; $\widetilde{\eta}$: SD; $\breve{\eta}$: modified CLS; $\overline{\eta}$: corrected SD.

Table 3.3: Biases ($\lambda = 5.0$).

T	α	Estimator of α				Estimator of λ		
		$\widehat{\alpha}$	$\widetilde{\alpha}$	$\breve{\alpha}$	$\overline{\alpha}$	$\widehat{\lambda}$	$\widetilde{\lambda}$	$\breve{\lambda}$
15	0.35	−0.0475	−0.0070	−0.0109	−0.0040	0.3564	0.0427	0.0712
	0.50	−0.0554	−0.0078	−0.0141	−0.0045	0.5183	0.0438	0.1093
	0.65	−0.0368	−0.0058	−0.0115	−0.0028	0.4573	0.0227	0.1169
	0.80	−0.0108	−0.0016	−0.0203	0.0006	0.2459	0.0050	0.4027
30	0.35	−0.0299	−0.0009	−0.0064	0.0006	0.2107	−0.0144	0.0297
	0.50	−0.0281	−0.0044	−0.0019	−0.0028	0.2514	0.0174	−0.0062
	0.65	−0.0172	−0.0018	−0.0045	−0.0003	0.2298	0.0108	0.0519
	0.80	−0.0081	−0.0018	−0.0134	−0.0008	0.1946	0.0377	0.3373
45	0.35	−0.0213	−0.0006	−0.0037	0.0004	0.1501	−0.0090	0.0141
	0.50	−0.0182	−0.0014	−0.0038	−0.0003	0.1737	0.0052	0.0328
	0.65	−0.0107	−0.0012	−0.0034	-0.0001	0.1507	0.0164	0.0449
	0.80	−0.0052	−0.0004	−0.0066	0.0003	0.0994	−0.0180	0.1308
60	0.35	−0.0146	−0.0005	−0.0004	0.0003	0.1130	0.0040	0.0039
	0.50	−0.0120	−0.0007	−0.0016	0.0001	0.1164	0.0022	0.0130
	0.65	−0.0082	−0.0009	−0.0027	−0.0002	0.1064	0.0028	0.0279
	0.80	−0.0043	−0.0008	−0.0042	−0.0002	0.0916	0.0032	0.0882

$\widehat{\eta}$: CML; $\widetilde{\eta}$: SD; $\breve{\eta}$: modified CLS; $\overline{\eta}$: corrected SD.

bias, showing a large superiority. Observe that $\breve{\alpha}$ has indeed a smaller bias than CML; however, even the uncorrected SD will be preferable. Also, the table indicates that although the modified CLS for λ has a substantially smaller bias than CML, there may be situations where the bias will be relatively large, particularly for small sample size.

Table 3.4: Integrated relative absolute bias.

T	Estimator of α				Estimator of λ		
	$\widehat{\alpha}$	$\widetilde{\alpha}$	$\breve{\alpha}$	$\overline{\alpha}$	$\widehat{\lambda}$	$\widetilde{\lambda}$	$\breve{\lambda}$
15	1.036	0.2122	0.3149	0.1007	0.9612	0.0800	0.4538
30	0.5794	0.0966	0.1388	0.0350	0.5198	0.0490	0.2345
45	0.3983	0.0571	0.0890	0.0139	0.3356	0.0310	0.0963
60	0.2896	0.0420	0.0580	0.0065	0.2593	0.0184	0.0658

$\widehat{\eta}$: CML; $\widetilde{\eta}$: SD; $\breve{\eta}$: modified CLS; $\overline{\eta}$: corrected SD.

Tables 3.5, 3.6 and 3.7 give, for $\lambda = 1$, $\lambda = 3$ and $\lambda = 5$, respectively, the estimated mean square errors for our four estimators of α and three estimators of λ. Table 3.8 summarizes our results, introducing a quantity that we call here the total relative deviation (TRD). Let η be an unknown parameter to be estimated and suppose that η^* is any estimator of η. Suppose that we estimate, by simulation, the mean square error of η^* as estimator of η for N different situations. Let η_j be the value of η for situation j (observe that some of the η_j will be equal) and let m_j be the estimated mean square error of η^* for this situation. We define the TRD as

$$TRD = \left(\sum_{j=1}^{N} \frac{m_j}{\eta_j^2} \right)^{1/2}.$$

From the Table 3.8, we conclude that the CML estimators achieve, in general, the smallest mean square errors, while the modified CLS achieve the largest. The mean square errors do not present much variation among the four estimators of α. However, the modified CLS estimator for λ has a considerably larger mean square error than the other two, in particular, for small sample sizes. On the other hand, the mean square errors for $\widehat{\lambda}$ and $\widetilde{\lambda}$ are very similar.

In summary, our proposed estimator for α, the corrected SD, has extremely good performance in terms of bias and its mean square error is almost as good as the mean square error of the CML estimator. Additionally, the SD estimator for λ, which is unbiased, has also a mean square error that is almost as good as the mean square error of the CML estimator. We therefore recommend the use of $\overline{\alpha}$ and $\widetilde{\lambda}$ as estimators of the parameters of an INAR(1) process if bias is of great concern, for example, if the sample size is very small.

3.5 Empirical illustration

In this section, we apply the estimation methods considered in the previous section to a real situation. The data set consists of monthly counts of claimants collecting Short Term Wage Loss Benefit (STWLB) for burn related injuries received in the workplace. In the selected data set all the claimants are male, with ages between 35 and 54, work in the logging industry and reported their claim to the Richmond, BC, service delivery location. The data set contains 120 observations starting in January 1985 and ending in December 1994. The claims counts are low with slightly higher values occurring

Table 3.5: Mean squared errors ($\lambda = 1.0$).

T	α	Estimator of α				Estimator of λ		
		$\widehat{\alpha}$	$\widetilde{\alpha}$	$\breve{\alpha}$	$\overline{\alpha}$	$\widehat{\lambda}$	$\widetilde{\lambda}$	$\breve{\lambda}$
15	0.35	0.0539	0.0831	0.1075	0.0910	0.2044	0.2381	0.3189
	0.50	0.0537	0.0519	0.1000	0.0553	0.2768	0.2276	0.4770
	0.65	0.0354	0.0262	0.0974	0.0271	0.2851	0.2151	0.8848
	0.80	0.0119	0.0085	0.0951	0.0085	0.2652	0.2104	2.6044
30	0.35	0.0305	0.0411	0.0417	0.0428	0.0999	0.1184	0.1230
	0.50	0.0244	0.0254	0.0375	0.0261	0.1076	0.1122	0.1636
	0.65	0.0138	0.0130	0.0339	0.0132	0.1144	0.1094	0.2984
	0.80	0.0043	0.0041	0.0285	0.0040	0.1056	0.1017	0.7737
45	0.35	0.0200	0.0276	0.0252	0.0284	0.0626	0.0779	0.0746
	0.50	0.0156	0.0169	0.0234	0.0172	0.0704	0.0743	0.1055
	0.65	0.0076	0.0079	0.0193	0.0079	0.0676	0.0698	0.1697
	0.80	0.0027	0.0027	0.0148	0.0027	0.0699	0.0719	0.3765
60	0.35	0.0147	0.0205	0.0184	0.0209	0.0450	0.0568	0.0540
	0.50	0.0107	0.0122	0.0168	0.0124	0.0488	0.0558	0.0743
	0.65	0.0057	0.0061	0.0133	0.0061	0.0489	0.0529	0.1130
	0.80	0.0019	0.0019	0.0100	0.0019	0.0477	0.0493	0.2603

$\widehat{\eta}$: CML; $\widetilde{\eta}$: SD; $\breve{\eta}$: modified CLS; $\overline{\eta}$: corrected SD.

Table 3.6: Mean squared errors ($\lambda = 3.0$).

T	α	Estimator of α				Estimator of λ		
		$\widehat{\alpha}$	$\widetilde{\alpha}$	$\breve{\alpha}$	$\overline{\alpha}$	$\widehat{\lambda}$	$\widetilde{\lambda}$	$\breve{\lambda}$
15	0.35	0.0540	0.0777	0.1004	0.0798	1.3895	1.8153	2.3570
	0.50	0.0518	0.0444	0.1012	0.0453	2.0108	1.6704	3.8334
	0.65	0.0315	0.0209	0.0958	0.0212	2.2808	1.5729	7.3002
	0.80	0.0091	0.0068	0.0979	0.0068	1.9429	1.4954	22.613
30	0.35	0.0308	0.0382	0.0408	0.0387	0.7376	0.8682	0.9462
	0.50	0.0222	0.0219	0.0359	0.0221	0.8473	0.8191	1.3754
	0.65	0.0115	0.0103	0.0324	0.0103	0.8581	0.7836	2.4237
	0.80	0.0037	0.0035	0.0274	0.0035	0.8418	0.7776	6.3394
45	0.35	0.0193	0.0241	0.0244	0.0243	0.4647	0.5569	0.5731
	0.50	0.0139	0.0145	0.0219	0.0146	0.5296	0.5566	0.8192
	0.65	0.0069	0.0067	0.0183	0.0067	0.5157	0.5076	1.3608
	0.80	0.0022	0.0022	0.0142	0.0022	0.5078	0.4936	3.2468
60	0.35	0.0140	0.0183	0.0172	0.0184	0.3323	0.4237	0.4008
	0.50	0.0095	0.0109	0.0156	0.0110	0.3716	0.4281	0.5935
	0.65	0.0047	0.0051	0.0124	0.0051	0.3620	0.3846	0.9386
	0.80	0.0017	0.0017	0.0010	0.0017	0.3800	0.3780	2.2587

$\widehat{\eta}$: CML; $\widetilde{\eta}$: SD; $\breve{\eta}$: modified CLS; $\overline{\eta}$: corrected SD.

Table 3.7: Mean squared errors ($\lambda = 5.0$).

T	α	Estimator of α				Estimator of λ		
		$\widehat{\alpha}$	$\widetilde{\alpha}$	$\breve{\alpha}$	$\overline{\alpha}$	$\widehat{\lambda}$	$\widetilde{\lambda}$	$\breve{\lambda}$
15	0.35	0.0542	0.0738	0.1048	0.0750	3.6019	4.6195	6.5296
	0.50	0.0520	0.0451	0.1013	0.0456	5.4106	4.6558	10.274
	0.65	0.0313	0.0214	0.0966	0.0215	6.2185	4.2810	20.117
	0.80	0.0088	0.0065	0.0940	0.0065	5.5988	4.0835	60.031
30	0.35	0.0292	0.0376	0.0385	0.0380	1.8318	2.3064	2.3751
	0.50	0.0222	0.0210	0.0356	0.0212	2.2693	2.1384	3.6536
	0.65	0.0112	0.0098	0.0324	0.0098	2.3113	2.0616	6.6565
	0.80	0.0035	0.0032	0.0281	0.0032	2.1882	2.0389	17.963
45	0.35	0.0196	0.0247	0.0240	0.0248	1.2338	1.5354	1.4956
	0.50	0.0137	0.0139	0.0216	0.0140	1.4277	1.4567	2.2452
	0.65	0.0065	0.0064	0.0188	0.0064	1.3622	1.3498	3.9142
	0.80	0.0022	0.0021	0.0151	0.0021	1.3621	1.3047	9.3933
60	0.35	0.0137	0.0186	0.0169	0.0187	0.8850	1.1566	1.0707
	0.50	0.0091	0.0102	0.0149	0.0102	0.9487	1.0499	1.5441
	0.65	0.0046	0.0048	0.0129	0.0048	0.9615	0.9965	2.6604
	0.80	0.0016	0.0016	0.0099	0.0016	1.0034	0.9977	6.2214

$\widehat{\eta}$: CML; $\widetilde{\eta}$: SD; $\breve{\eta}$: modified CLS; $\overline{\eta}$: corrected SD.

Table 3.8: Total relative deviation.

T	Estimator of α				Estimator of λ		
	$\widehat{\alpha}$	$\widetilde{\alpha}$	$\breve{\alpha}$	$\overline{\alpha}$	$\widehat{\lambda}$	$\widetilde{\lambda}$	$\breve{\lambda}$
15	1.494	1.636	2.213	1.670	1.647	1.525	3.490
30	1.057	1.150	1.338	1.161	1.066	1.070	1.954
45	0.8452	0.9303	1.035	0.9363	0.8427	0.8688	1.440
60	0.7120	0.8049	0.8613	0.8086	0.7092	0.7498	1.195

$\widehat{\eta}$: CML; $\widetilde{\eta}$: SD; $\breve{\eta}$: modified CLS; $\overline{\eta}$: corrected SD.

between January 1990 and December 1992, which could be caused by a single claimant with a severe or recurring dislocation. These data were previously studied by Freeland (1998) and are listed in Table 3.9. The required numerical evaluations are here implemented using the R software.

Table 3.9: Burns claims.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1985	0	0	1	1	1	1	1	1	0	1	1	2
1986	0	0	0	0	1	0	0	1	1	1	1	1
1987	0	0	0	1	0	0	0	0	0	0	2	1
1988	0	0	2	2	2	2	1	0	0	1	1	1
1989	1	1	2	2	2	2	2	3	1	2	2	1
1990	1	0	1	3	1	1	1	1	2	1	1	2
1991	2	1	1	3	4	3	2	1	1	0	0	0
1992	1	2	0	0	0	0	0	0	0	1	1	1
1993	1	1	1	1	0	0	0	1	1	1	1	1
1994	0	0	0	0	0	1	0	0	1	2	3	2

Table 3.10 provides some descriptive statistics. We note that the ratio between the sample variance and sample mean is 0.836, and then, the data seem to be equidispersed. The equidispersion test (Rao and Chakravarti, 1956) did not reject the null hypothesis of equidispersion, the p-value for the test being 0.90. Although the test supposes independent observations, the very high p-value that was obtained indicates that the hypothesis of equidispersion may be a reasonable one. Consequently, a Poisson marginal distribution seems to be appropriate. The series and its sample autocorrelation function are displayed in Figure 3.1.

Table 3.10: Descriptive statistics.

Minimum	Median	Mean	Variance	$\widehat{\rho}(1)$	Maximum
0	1	0.917	0.766	0.583	4

Analyzing Figure 3.1, we conclude that a first order autoregressive model may be appropriate for the current data series, because of the geometric decrease in the sample autocorrelations (as the lag increases) and the clear cut-off after lag 1 in the partial autocorrelations. Furthermore, the behavior of the series indicates that it may be mean stationary.

Table 3.11 presents, for the first 30, 45 and 60 observations, the conditional maximum likelihood estimates, the SD estimates, the modified CLS estimates and the corrected SD estimates, with the corresponding standard errors in parenthesis. Three goodness-of-fit statistics are also presented: the RMS (root mean square error), MAE (mean absolute error) and AME (absolute median error). Those statistics are defined as follows. For $t = 2, \ldots, T$, consider the expected value of the observation at the previous time, $\mathrm{E}[Y_t|Y_{t-1}] = \alpha Y_{t-1} + \lambda$. If α^* and λ^* are any estimators of α and λ, then, for $t = 2, \ldots, T$, we define the corresponding absolute deviations as $a_t^* = |Y_t - \alpha^* Y_{t-1} - \lambda^*|$. The RMS is obtained as the square root of the average value of $(a_t^*)^2$, the MAE is obtained as the average value of a_t^* and the AME is obtained as the median of the values of a_t^*. The RMS has also been considered by Zhu and Joe (2006)

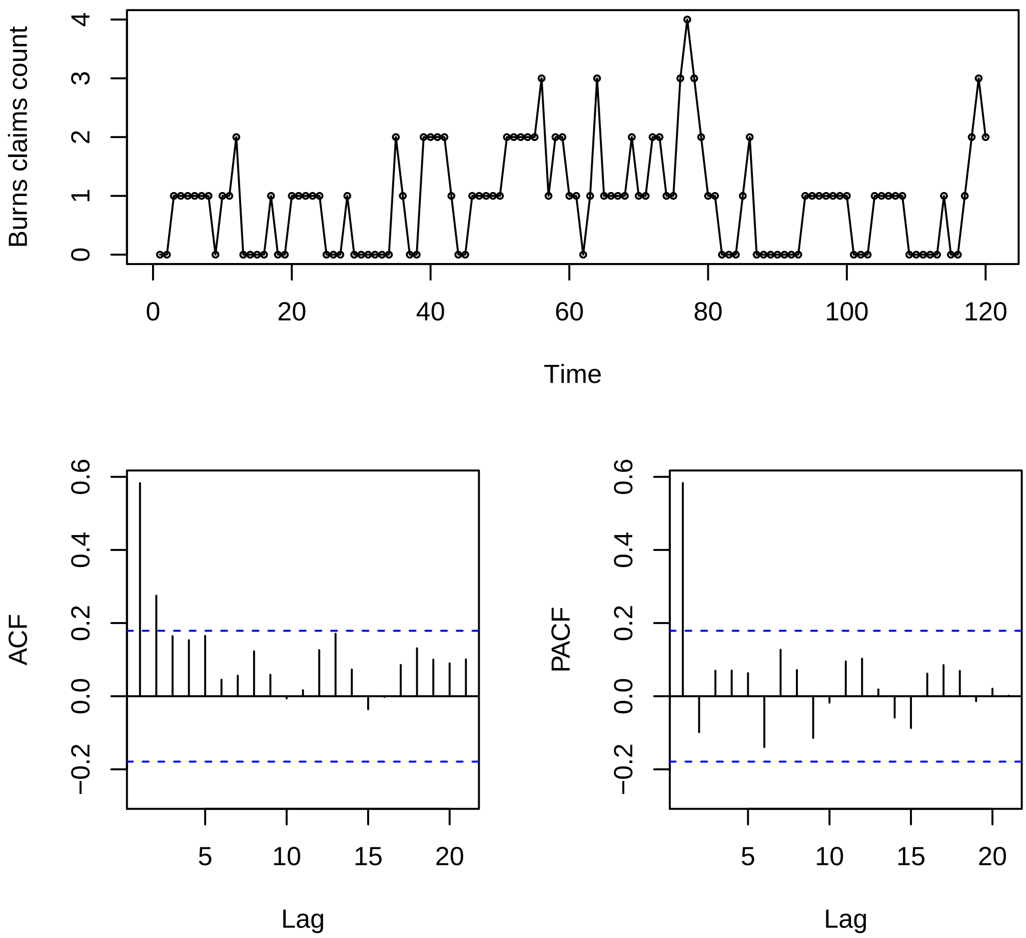


Figure 3.1: The time series plot, autocorrelation and partial autocorrelation functions of the burns claims series from 1985 to 1994.

for comparing different fittings in INAR(1) models. Since the Fisher's information matrix is not available, the standard errors are obtained as the square roots of the elements in the diagonal of the inverse of the negative of the Hessian of the conditional log-likelihood calculated at the conditional maximum likelihood estimates.

Table 3.11: Estimates of the parameters (standard errors in parentheses), RMS, MAE and AME for counts of burns claims.

T	Estimator	Parameter				RMS	MAE	AME
		α		λ				
30	$(\widehat{\alpha},\widehat{\lambda})$	0.517	(0.176)	0.283	(0.124)	0.568	0.468	0.283
	$(\widetilde{\alpha},\widetilde{\lambda})$	0.574	(0.168)	0.241	(0.112)	0.561	0.525	0.433
	$(\breve{\alpha},\breve{\lambda})$	0.287	(0.205)	0.418	(0.156)	0.562	0.528	0.414
	$(\overline{\alpha},\widetilde{\lambda})$	0.608	(0.156)	0.241	(0.111)	0.584	0.456	0.241
45	$(\widehat{\alpha},\widehat{\lambda})$	0.524	(0.133)	0.314	(0.105)	0.650	0.523	0.314
	$(\widetilde{\alpha},\widetilde{\lambda})$	0.542	(0.139)	0.296	(0.105)	0.708	0.629	0.644
	$(\breve{\alpha},\breve{\lambda})$	0.459	(0.161)	0.357	(0.125)	0.708	0.630	0.659
	$(\overline{\alpha},\widetilde{\lambda})$	0.560	(0.134)	0.296	(0.105)	0.654	0.518	0.295
60	$(\widehat{\alpha},\widehat{\lambda})$	0.658	(0.088)	0.318	(0.090)	0.679	0.513	0.365
	$(\widetilde{\alpha},\widetilde{\lambda})$	0.664	(0.091)	0.297	(0.090)	0.798	0.643	0.883
	$(\breve{\alpha},\breve{\lambda})$	0.577	(0.125)	0.390	(0.126)	0.797	0.633	0.921
	$(\overline{\alpha},\widetilde{\lambda})$	0.677	(0.088)	0.297	(0.090)	0.682	0.511	0.350

From Table 3.11, it can be seen that the modified CLS behaves quite differently from the other three, in particular for a small number of observations. It is also the estimator with the largest standard errors and worst goodness-of-fit statistics. The corrected SD estimator, on the other hand, produces the estimates with smallest standard errors in all situations. Also, in terms of root mean square error, all estimators are competitive for $T = 30$, while, for 45 and 60 observations, the CML and corrected SD estimators are competitive and superior. In terms of mean absolute error, we conclude that the CML and corrected SD estimators are also competitive and superior to the other estimators, the corrected SD estimators being slightly better. In terms of absolute median error, CML and corrected SD estimators are clearly superior to the other estimators, the corrected SD estimator being the best estimator. Hence, we conclude that, in terms of fitting, the corrected SD estimator is the best of all four alternatives.

We can also compare the estimators in terms of predictive power. Let again α^* and λ^* be any estimators of α and λ, obtained from T observations. Then, at any given time T, we can forecast the next observation as

$$Y^*_{T+1} = \langle\, \alpha^* Y_T + \lambda^* \,\rangle\,,$$

where $\langle\cdot\rangle$ represents the nearest integer.

Table 3.12: The absolute deviation $|Y^*_{T+1} - Y_{T+1}|$ for counts of burns claims.

T	$(\widehat{\alpha},\widehat{\lambda})$	$(\widetilde{\alpha},\widetilde{\lambda})$	$(\breve{\alpha},\check{\lambda})$	$(\overline{\alpha},\widetilde{\lambda})$
45	1	1	1	1
46	0	0	0	0
47	0	0	0	0
48	0	0	0	0
49	0	0	0	0
50	1	1	1	1
51	1	1	1	0
52	0	0	1	0
53	0	0	1	0
54	0	0	1	0
$\sum_{T=45}^{54} \lvert Y^*_{T+1} - Y_{T+1}\rvert$	3	3	6	2

Table 3.12 shows that all four estimators produce a correct forecast or a small forecast error in all the time interval here studied ($T = 45,\ldots,54$). This confirms that the INAR(1) process is a good model for our data, at least in this interval. We observe that the modified CLS provides the worst forecasting performance, since it is the only one producing a wrong forecast for $T = 52, 53, 54$. On the other hand, the corrected SD provides the best performance, being the only estimator that produces a correct forecast for $T = 51$. This is an indication that our proposed corrected SD estimator can be indeed a good choice for forecasting purposes.

3.6 Concluding remarks

In this chapter, we have examined a wide range of estimators for the two parameters of the Poisson INAR(1) process. We have derived a closed-form expression for the second order bias of the SD estimator of the parameter α that index the Poisson INAR(1) process. Such expression was used to construct a bias-corrected estimator. The bias of the proposed estimator has order $O(T^{-2})$. Furthermore, asymptotic properties of the estimators have been studied. The numerical evidence shows that the estimator we propose has good finite-sample behavior, even when the sample size is small. Furthermore, while most estimates, like CML, are not so simple to obtain, the SD and corrected SD estimates are indeed very simple to obtain from time series data. We, therefore, strongly recommend practitioners to use the SD estimator for λ and our proposed corrected SD estimator for α when modeling time series of counts by a Poisson INAR(1) process. Finally, we have applied our proposed estimator to a real time series data and observed that our proposed estimators can produce very good forecasts.

3.7 Appendix

Proof of Proposition 2

In Proposition 4.3.1 of Freeland (1998), it is shown that the stationary Poisson INAR(1) process is α-mixing with exponentially decreasing weights. Consider the bivariate process $\boldsymbol{\xi}_t$ defined by $\boldsymbol{\xi}_t = (Y_t - \mu, (Y_t - Y_{t-1})^2 - 2\lambda)$. Since $\boldsymbol{\xi}_t$ is a function of finitely many random variables from Y_t, $\boldsymbol{\xi}_t$ is also α-mixing with exponentially decreasing weights. Furthermore, $\mathrm{E}[(Y_t - Y_{t-1})^{24}] < \infty$ (Weiß, 2012)..

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

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

ξ_t^h representing the h-th component of $\boldsymbol{\xi}_t$.

Then,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mu\frac{1+\alpha}{1-\alpha} & 2\,\mu \\ 2\,\mu & 4\,\lambda\left(1+\lambda\frac{3+\alpha}{1+\alpha}\right) \end{pmatrix}.$$

Now, suppose $f = (f_1, f_2)^\top$ is a mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$, where each f_i is a function from $\mathbb{R}^2$ to $\mathbb{R}$ that is differentiable at a given point $\boldsymbol{\eta}$. Let $\mathbf{D}$ be the 2×2 Jacobian matrix of f with respect to $\boldsymbol{\eta}$. Then, if $\{\boldsymbol{X}_T; T \geq 1\}$ is a random bivariate sequence such that $\sqrt{T}\,(\boldsymbol{X}_T - \boldsymbol{\eta}) \xrightarrow{d} \mathrm{N}\,(0, \boldsymbol{\Sigma})$, then, according to the delta method, we have $\sqrt{T}(f(\boldsymbol{X}_T) - f(\boldsymbol{\eta})) \xrightarrow{d} \mathrm{N}\left(0, \mathrm{D}\boldsymbol{\Sigma}\mathrm{D}^\top\right)$.

For $f_1(x_1, x_2) = 1 - x_2/2\,x_1$, $f_2(x_1, x_2) = x_2/2$ and $\boldsymbol{\eta} = (\lambda/(1-\alpha), 2\lambda)^\top$, we obtain

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

Let (α^*, λ^*) be given by $\left(f\left(\frac{1}{T}\sum_{t=1}^{T}Y_t, \frac{1}{T}\sum_{t=1}^{T}(Y_t - Y_{t-1})^2\right)\right)$. Then, according to the delta method, we have

$$\sqrt{T}\left((\alpha^*, \lambda^*)^\top - (\alpha, \lambda)^\top\right) \xrightarrow{d} \mathrm{N}\left(0, \mathrm{D}\boldsymbol{\Sigma}\mathrm{D}^\top\right),$$

where

$$\mathbf{D} = \begin{pmatrix} \frac{\lambda}{\mu^2} & -\frac{1}{2\,\mu} \\ 0 & \frac{1}{2} \end{pmatrix},$$

so that

$$\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^\top = \begin{pmatrix} \frac{\alpha(1-\alpha)^2}{\lambda} + (1-\alpha)^2\frac{3+\alpha}{1+\alpha} & -\lambda(1-\alpha)\frac{3+\alpha}{1+\alpha} \\ -\lambda(1-\alpha)\frac{3+\alpha}{1+\alpha} & \lambda\left(1+\lambda\frac{3+\alpha}{1+\alpha}\right) \end{pmatrix}.$$

Finally, observe that the SD estimators $\widetilde{\alpha}$ and $\widetilde{\lambda}$ of α and λ are related to α^* and λ^* as

$$\begin{aligned} \widetilde{\alpha} &= \alpha^* - \frac{1}{T-1}\left(\frac{\lambda^*}{\overline{Y}} + \frac{(Y_1 - Y_0)^2}{2\overline{Y}}\right), \\ \widetilde{\lambda} &= \lambda^* + \frac{1}{T-1}\left(\lambda^* - \frac{(Y_1 - Y_0)^2}{2}\right), \end{aligned}$$

and the result is proved.

Proof of Proposition 3

From Freeland and McCabe (2005), we have

$$\sqrt{T}(\breve{\alpha}-\alpha) \xrightarrow{d} \mathrm{N}\left(0, \frac{\alpha(1-\alpha)^2}{\lambda} + (1-\alpha)(1+\alpha)\right)$$

and

$$\sqrt{T}(\breve{\lambda}-\lambda) \xrightarrow{d} \mathrm{N}\left(0, \lambda\left(1+\frac{1+\alpha}{1-\alpha}\lambda\right)\right).$$

Thus,

$$\begin{aligned}
\sqrt{T}(\check{\alpha}-\alpha) &= \sqrt{T}\left(\frac{T\breve{\alpha}}{T-3}+\frac{1}{T-3}-\alpha\right) = \sqrt{T}\left(\frac{T\breve{\alpha}}{T-3}+\frac{T\alpha}{T-3}-\frac{T\alpha}{T-3}-\alpha\right)+\frac{\sqrt{T}}{T-3} \\
&= \frac{T}{T-3}\sqrt{T}\,(\breve{\alpha}-\alpha)+\alpha\sqrt{T}\left(\frac{T}{T-3}-1+\frac{1}{\alpha(T-3)}\right) \\
&\xrightarrow{d} \mathrm{N}\left(0, \frac{\alpha(1-\alpha)^2}{\lambda}+(1-\alpha)(1+\alpha)\right).
\end{aligned}$$

Now consider

$$\breve{\lambda} = \frac{1}{T-1}\left(\sum_{t=2}^{T} Y_t - \breve{\alpha}\sum_{t=2}^{T} Y_{t-1}\right) \quad \text{and} \quad \check{\lambda} = \frac{1}{T-1}\left(\sum_{t=2}^{T} Y_t - \check{\alpha}\sum_{t=2}^{T} Y_{t-1}\right).$$

Then,

$$\check{\lambda}-\breve{\lambda} = \overline{Y}_1(\breve{\alpha}-\check{\alpha}) = \overline{Y}_1\left(\breve{\alpha}-\frac{T\breve{\alpha}+1}{T-3}\right) = \overline{Y}_1\left(\frac{1-3\breve{\alpha}}{T-3}\right),$$

where $\overline{Y}_1 = \sum_{t=2}^{T} Y_{t-1}/(T-1)$. Thus,

$$\begin{aligned}
\sqrt{T}(\check{\lambda}-\lambda) &= \sqrt{T}(\breve{\lambda}-\lambda)+\overline{Y}_1\frac{\sqrt{T}}{T-3}(1-3\breve{\alpha}) \\
&\xrightarrow{d} \mathrm{N}\left(0, \lambda\left(1+\frac{1+\alpha}{1-\alpha}\lambda\right)\right).
\end{aligned}$$

References

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

Billingsley, P. *Probability and Measure*, John Wiley & Sons, New York, 1979.

Brännäs, K., Hellström, J. and Nordström, J. (2002) . A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting*, **18**, 19–30.

Franke,J. and Seligmann, T. (1993). Conditional maximum-likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In *Developments in Time Series* (ed. T. Subba Rao). London: Chapman & Hall, 310–330.

Freeland, R.K. *Statistical analysis of discrete time series with applications to the analysis of workers compensation claims data*, Ph.D. thesis, University of British Columbia, Canada, 1998.

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

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

Hellström, J. (2001). Unit root testing in integer-valued AR(1) models. *Economics Letters*, **70**, 9–14.

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

Kendall, M.G. (1954). Note on the bias in the estimation of autocorrelation. *Biometrika*, **41**, 207–214.

Mariott, F.H.C. and Pope, J.A. (1954). Bias in the estimation of autocorrelations. *Biometrika*, **41**, 393–403.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, **21**, 645–650.

Park, Y. and Oh, C.W. (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals. *Statistical Papers*, **38**, 287–302.

Orcutt, G.H. and Winokur, H.S. (1969). First order autoregression: inference, estimation, and prediction. *Econometrica*, **37**, 1–14.

Rao, C.R. and Chakravarti, I.M. (1956). Some small sample tests of significance for a Poisson distribution, *Biometrics*, **12**, 264–282.

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

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

Weiß, C.H. (2011). Simultaneous confidence regions for the parameters of a Poisson INAR(1) model. *Statistical Methodology*, **8**, 517–527.

Weiß, C.H. (2012). Process capability analysis for serially dependent processes of Poisson counts. *Journal of Statistical Computation and Simulation*, **82**, 383–404.

Zhu, R. and Joe, H. (2006). Modelling count data time series with Markov processes based on binomial thinnning. *Journal of Time Series Analysis*, **27**, 725–738.

CHAPTER 4

The effects of additive outliers in INAR(1) processes and robust estimation

Resumo

Neste capítulo, propomos duas metodologias de estimação para os parâmetros do processo Poisson INAR(1), as quais são robustas na presença de observações atípicas ou outliers. O método de estimação é uma variação das equações de Yule-Walker. Em particular, esse estudo propõe a estimação da função de autocorrelação do processo Poisson INAR(1) através do coeficiente de correlação de Spearman e do coeficiente de correlação Gaussiano. As estimativas dessas funções são substituídas nas equações de Yule-Walker fornecendo as estimativas para os parâmetros do modelo Poisson INAR(1). Resultados de Monte Carlo mostraram que, em geral, os estimadores propostos para os parâmetros do processo Poisson INAR(1) são robustos na presença de outliers aditivos.

Palavras-chave: Outliers aditivos. Processos Poisson INAR(1). Estimador diferença de quadrados. robustês.

Abstract

In this chapter, methods based on ranks for estimating the parameters of the first-order integer-valued autoregressive model in the presence of additive outliers are proposed. In particular, we use the robust sample autocorrelations based on ranks to obtain estimators for the parameters of the Poisson INAR(1) process. The effects of additive outliers on the estimates of parameters of integer-valued time series are examined. Some numerical results of the estimators are presented with a discussion of the obtained results. The proposed methods are applied to a dataset concerning the number of different IP addresses accessing the server of the pages of the Department of Statistics of the University of Würzburg. The results presented here give motivation to use the methodology in practical situations in which Poisson INAR(1) processes contain additive outliers.

Keywords: Additive outliers. Poisson INAR(1) processes. Robustness. Squared difference estimator.

4.1 Introduction

Time series of counts occur in many contexts, often as counts of events, objects or individuals in consecutive intervals or at consecutive points in time. Integer-valued time series data or count data arise naturally in many areas including epidemiology (Zeger and Qaqish, 1988), medicine (Franke and Seligmann, 1993), finance (Brännäs et al., 2002), economics (Freeland and McCabe, 2004), health science (Kachour and Yao, 2009) and others. In recent years, several models for the analysis of time series of counts have been developed. For a review of models for analysis of time series of counts, see Weiß (2008).

In practical situations, it is quite common to have samples that have discrepant observations and they usually occur as a result of measurement errors, an influence of exogenous variables, an unexpected phenomenon among other situations. Such atypical observations are usually called outliers (Fox, 1972). There are different classes of outliers which have quite different impacts on an estimate (Denby and Martin, 1979). However, the most common types considered in the literature are: innovation outliers (IO), which produces an effect in all subsequent observations, additive outliers (AO), which affects only the level of the contaminated observation.

Several classes of robust estimators for ARMA models have been proposed (Martin and Yohai, 1985). Recently, the influence of outliers in time series model estimation has been the focus of much research (Fajardo et al., 2009; Sarnaglia et al., 2010). For a review of robust estimation for ARMA models, see Martin and Yohai (1985). However, the study of robustness in time series of counts has not received much attention so far in the literature. Barczy et al., (2010, 2012) consider conditional least squares (CLS) estimation of the parameters of an INAR(1) model contaminated, at known time periods, with innovational and additive outliers, respectively. The problem of detecting outliers in Poisson INAR(1) processes has been investigated by Silva and Pereira (2012). The motivation for studying integer-valued time series with atypical observations can be the fact they are quite common in many fields of application. Hence, it is an interesting research area to work with. Furthermore, robust procedures for estimating the parameters of the Poisson INAR(1) process has not yet been addressed in the literature. This paper aims to give a contribution in this direction. In this paper, robust estimators for the parameters of the Poisson INAR(1) process in the presence of additive outliers are obtained by replacing the classical autocovariance in the Yule-Walker equations by the robust autocovariance based on ranks.

The rest of the chapter unfolds as follows. In Section 4.2, the Poisson INAR(1) process is introduced, some of its basic properties are outlined and estimation methods for the model parameters are described. The impact of outliers on the correlation structure and estimation of the Poisson INAR(1) process are derived in Section 4.3. In Section 4.4, we describe robust procedures for estimating the autocorrelation function (ACF) and robust estimators for the parameters of the Poisson INAR(1) are proposed. Section 4.5 discusses some simulation results for the estimation methods. An application to real data is performed in Section 6. Some concluding remarks are addressed in Section 4.7.

4.2 Background: The Poisson INAR(1) process

In this section we provide a brief background of the Poisson INAR(1) process and the estimation of the unknown parameters.

4.2.1 The Poisson INAR(1) Model

The INAR(1) model is based on the probabilistic operation of binomial thinning. Consider the thinning operation " $\circ$ ", introduced by Steutel and Van Harn (1979) and let $\alpha \in [0, 1]$. If X is any nonnegative integer-valued random variable, $\alpha \circ X$ is defined as $\alpha \circ X = \sum_{j=1}^{X} Z_j$, where $\{Z_j\}_{j=1}^{X}$ are independent and identically distributed (i.i.d.) random variables, independent of X, with $\Pr(Z_j = 1) = 1 - \Pr(Z_j = 0) = \alpha$, i.e., $\{Z_j\}_{j=1}^{X}$ is an i.i.d. Bernoulli random sequence.

A discrete-time integer-valued stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$ is said to be a first-order integer-valued autoregressive (INAR(1)) process if it satisfies the following equation:

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t,\ t \in \mathbb{Z}, \tag{4.1}$$

where $\alpha \in [0,1]$ and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed non-negative integer-valued random variables. ϵ_t and Y_{t-i} being independent for all $i \geq 1$. It is also assumed that the Bernoulli variables that define $\alpha \circ Y_{t-1}$, that is, the Bernoulli variables from which Y_t are obtained, are independent of the Bernoulli variables from which other values of the series are calculated. As for the first order autoregressive process with normally distributed innovations, the conditional expectation of Y_t is linear in Y_{t-1}.

In this paper, it is assumed that $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence of Poisson distributed variables with mean $\lambda \in \mathbb{R}_+$ and that, for all t, this sequence is mutually independent of all Bernoulli random variables that define $\alpha \circ Y_t$. Also, Y_t has Poisson distribution with parameter $\lambda/(1-\alpha)$, i.e., $Y_t \sim \mathcal{P}o(\lambda/(1-\alpha))$ for all $t \in \mathbb{Z}$. Furthermore, the process $\{Y_t\}_{t\in\mathbb{Z}}$ satisfying (4.1) is second-order stationary if $0 \leq \alpha < 1$ (Du and Li, 1991). Many new results have been obtained in recent years for Poisson INAR(1) models. For example, Park and Oh (1997) studied the asymptotic properties of Yule-Walker estimators, Hellström (2001) focused on the testing of a unit root, Freeland and McCabe (2005) obtained asymptotic properties of conditional least squares estimators, Weiß (2011) proposed several asymptotic simultaneous confidence regions for the two parameters.

It is easy to verify that the ACF at lag k is given by

$$\mathrm{Corr}(Y_t, Y_{t+k}) = \rho_Y(k) = \alpha^k,\ k \geq 1, \tag{4.2}$$

which obviously is restricted to be positive. Equation (4.2) shows that the autocorrelation function, $\rho(k)$, decays exponentially with lag k as happens in the classical AR(1) model.

4.2.2 Estimation 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 given time series data. There are several ways to estimate the parameters of a Poisson INAR(1) process.

From a sample $Y_1, \ldots, Y_T$ of a stationary process $\{Y_t\}_{t\in\mathbb{Z}}$, the sample autocorrelation function is given by

$$\widehat{\rho}(k) = \frac{\widehat{\gamma}(k)}{\widehat{\gamma}(0)} = \frac{\sum\limits_{t=1}^{T-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2},$$

where $\overline{Y} = (1/T)\sum_{t=1}^{T} Y_t$ and $\widehat{\gamma}(k) = (1/T)\sum_{t=1}^{T-k}(Y_t - \overline{Y})(Y_{t-k} - \overline{Y})$ denote the sample mean and the sample autocovariance function, respectively. It is well-known that the estimators above are strongly consistent (Du and Li, 1991). The Yule-Walker (YW) estimators of α and λ are based upon the sample autocorrelation function $\widehat{\rho}(k)$, using that $\rho_Y(1) = \alpha$, and the first moment of Y_t, which is $\mathrm{E}(Y_t) = \lambda/(1-\alpha)$. They are given by

$$\widehat{\alpha}_{\mathrm{YW}} = \widehat{\rho}(1) = \frac{\sum\limits_{t=1}^{T-1}(Y_t - \overline{Y})(Y_{t+1} - \overline{Y})}{\sum_{t=1}^{T}(Y_t - \overline{Y})^2}, \qquad \widehat{\lambda}_{\mathrm{YW}} = (1 - \widehat{\alpha}_{\mathrm{YW}})\overline{Y}. \tag{4.3}$$

Al-Osh and Alzaid (1987) and Brännäs (1994) suggested the use of the conditional maximum likelihood (CML) estimation to obtain the estimates. However, the CML estimators do not have closed-form expressions. CLS estimation method was also considered in Freeland and McCabe (2005). Freeland and McCabe (2005) showed that the CLS estimators have the same asymptotic distribution as the YW estimators.

Weiß (2012) as well as Bourguignon and Vasconcellos (2014) suggested estimating the parameters of the process using the squared difference (SD) estimator given by

$$\widehat{\lambda}^{Y}_{\text{SD}} = \frac{1}{2(T-1)} \sum_{t=2}^{T} (Y_t - Y_{t-1})^2, \qquad \widehat{\alpha}^{Y}_{\text{SD}} = 1 - \frac{\widehat{\lambda}^{Y}_{\text{SD}}}{\overline{Y}}.$$

The alternative estimators discussed in the literature (YW, CML and CLS) perform much worse in terms of bias and mean square error than the SD estimator Bourguignon and Vasconcellos (2014). For a good discussion of estimation in the Poisson INAR(1) process, the reader is referred to (Jung et al., 2005) and Bourguignon and Vasconcellos (2014).

4.3 The impact of AO in the Poisson INAR(1) process

The model contaminated by additive outliers is defined here as

$$Z_t = Y_t + \omega\, \delta_t, \tag{4.4}$$

where $\omega \in \mathbb{N}$ is the magnitude of the outliers (fixed and unknown parameter), δ_t's are i.i.d. random variables, mutually independent of Y_t, with $\mathrm{P}(\delta_t = 1) = 1 - \mathrm{P}(\delta_t = 0) = p$, with $p \in (0,1)$, i.e., $\{\delta_t\}_{t\in\mathbb{N}}$ is an i.i.d. Bernoulli random sequence with mean p and variance $p(1-p)$. The product $T\,p$ is the expected number of outliers in the data. As previously discussed, outliers can affect the correlation structure of a time series and, consequently, the estimation of the model.

Proposition 3. *Suppose that $\{Z_t\}_{t\in\mathbb{Z}}$ follows process (4.4). The autocovariance function of $\{Z_t\}_{t\in\mathbb{Z}}$ is given by*

$$\gamma_Z(k) = \begin{cases} \gamma_Y(0) + \omega^2 p(1-p), & \text{if } k = 0; \\ \\ \gamma_Y(k), & \text{if } k \neq 0, \end{cases}$$

where $\gamma_Y(k) = E(Y_t\, Y_{t+k}) - E(Y_t)E(Y_{t+k})$ is the autocovariance function of $\{Y_t\}_{t\in\mathbb{Z}}$.

From this result, it can be seen that the AO increases the variance of $\{Z_t\}_{t\in\mathbb{N}}$, which provokes a reduction of the ACF of the process. In addition, for all positive k,

$$\rho_Z(k) \leq \rho_Y(k) \;\text{ and }\; \lim_{\omega\to\infty} \rho_Z(k) = 0. \tag{4.5}$$

Equation (4.5) shows that additive outliers introduce memory loss in the process $\{Z_t\}_{t\in\mathbb{Z}}$. This leads to estimates with significant negative bias, as shown in Section 4.5.

Proposition 4. *Let $Z_1, Z_2, \ldots, Z_T$ be generated from model (4.4) with one outlier with magnitude ω. It follows that:*

(i) The SD estimator for λ is given by

$$\widehat{\lambda}^{Z}_{\text{SD}} = \widehat{\lambda}^{Y}_{\text{SD}} + \frac{\omega^2 \sum_{t=2}^{T} \delta_t(\delta_t - \delta_{t-1})}{T-1};$$

(ii) The SD estimator for α is given by

$$\widehat{\alpha}^{Z}_{\mathrm{SD}} = 1 - \frac{\widehat{\lambda}^{Z}_{\mathrm{SD}}}{\overline{Z}};$$

where $\widehat{\lambda}^{Z}_{\mathrm{SD}}$ is the SD estimator of λ considering $\{Z_t\}_{t=1}^{T}$ and $\overline{Z} = (1/T)\sum_{t=1}^{T} Z_t = (1/T)\sum_{t=1}^{T}(Y_t + \omega\,\delta_t) = \overline{Y} + (\omega/T)\sum_{t=1}^{T}\delta_t$ is the sample mean of $\{Z_t\}_{t=1}^{T}$.

Corollary 2. *Assume that $\{Z_t\}_{t\in\mathbb{N}}$ is a stationary process with additive outliers as described in Equation (4.4). Then,*

(i) $\lim_{\omega\to 0}\widehat{\lambda}^{Z}_{SD} = \widehat{\lambda}^{Y}_{SD}$ *and* $\lim_{\omega\to 0}\widehat{\alpha}^{Z}_{SD} = \widehat{\alpha}^{Y}_{SD}$;

(iii) $\widehat{\lambda}^{Z}_{\mathrm{SD}} \geq \widehat{\lambda}^{Y}_{\mathrm{SD}}$;

(iv) $\mathrm{Bias}(\widehat{\lambda}^{Z}_{SD}) = \omega^2 p(1-p)$.

Corollary 2 shows that additive outliers introduce significant positive bias in the SD estimator for the λ parameter.

Propositions 3 and 4 show that the additive outliers can affect the statistical properties of the parameter estimates in Poisson INAR(1) processes. In this context, it is necessary to use robust methods for estimating time series models with outliers. This is the motivation of the next section.

4.4 Robust procedures to estimate parameters via nonparametric measures

The Spearman and Gaussian correlation coefficients are commonly used in nonparametric tests of independence between two variables (Hájek and Sidak, 1967). This chapter has no interest in testing the independence between variables but in estimating the degree of this dependence, this justifies the use of the coefficients in time series, i.e., here, the Spearman and Gaussian correlation coefficients, defined below, will be considered to estimate the ACF.

Let the pairs $(X_1, Y_1), \ldots, (X_T, Y_T)$ be a sample of size T of a bivariate random vector (X, Y), and $(R_1, S_1), \ldots, (R_T, S_T)$ the associated pairs of ranks from these observations. The Spearman rank correlation coefficient (Spearman, 1904) is defined by

$$\widehat{\rho}_{\mathrm{S}} = \frac{12\sum_{t=1}^{T}\left[(R_t - \frac{T+1}{2})(S_t - \frac{T+1}{2})\right]}{T(T^2-1)} = 1 - \frac{6\sum_{t=1}^{T} D_t^2}{T(T^2-1)},$$

where $D_t = S_t - R_t$. When there are ties in the observations $X's$ and/or $Y's$, the Spearman correlation coefficient is calculated as follows:

$$\widehat{\rho}_{\mathrm{S}} = \frac{T(T^2-1) - 6\sum_{t=1}^{T} D_t^2 - \frac{1}{2}\left\{\sum_{i=1}^{g}\left[v_i(v_i^2-1)\right] + \sum_{j=1}^{l}\left[u_j(u_j^2-1)\right]\right\}}{\left\{\left[T(T^2-1) - \sum_{i=1}^{g}\left[v_i(v_i^2-1)\right]\right]\left[T(T^2-1) - \sum_{j=1}^{l}\left[u_j(u_j^2-1)\right]\right]\right\}},$$

where g denotes the number of tied X groups, v_i is the size of tied X group i, l is the number of tied Y groups, and u_j is the size of tied Y group j.

For a series $\{Y_t\}_{t\in\mathbb{Z}}$ generated by any process, the $\widehat{\rho}_{\mathrm{S}}$ that is used to estimate $\rho(k)$ are defined as follows:

$$\widehat{\rho}_{\mathrm{S}}(k) = \frac{12\sum_{t=1}^{T}\left[(R_t - \frac{T+1}{2})(S_t - \frac{T+1}{2})\right]}{T(T^2-1)},$$

where (R_t, S_t) are the ranks of (Y_t, Y_{t+k}).

The Gaussian rank correlation (Hájek and Sidak, 1967) is given

$$\widehat{\rho}_{\mathrm{G}} = c_T \sum_{t=1}^{T} \Phi^{-1}\left(\frac{R_t}{T+1}\right) \Phi^{-1}\left(\frac{S_t}{T+1}\right), \tag{4.6}$$

where the constant $c_T = 1/\sum_{t=1}^{T}[\Phi^{-1}(t/(T+1))]^2$ and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. For the treatment of ties see Hájek and Sidak (1967). For additional properties of the Gaussian rank correlation estimator, see Boudt et al. (2012). Thus, for a series $\{Y_t\}_{t\in\mathbb{Z}}$ generated by any process, the $\widehat{\rho}_{\mathrm{G}}$ is used to estimate $\rho(k)$ using Equation (4.6) with (R_t, S_t) being the ranks of (Y_t, Y_{t+k}).

Thanks to the use of ranks, the Spearman and Gaussian correlations are resistant to small amounts of outliers in the data (Boudt et al., 2012). In this paper, robust estimators of the Poisson INAR(1) parameters are obtained by replacing the classical Yule-Walker estimator in Equation (4.3) by the $\widehat{\rho}_{\mathrm{S}}(k)$ and $\widehat{\rho}_{\mathrm{G}}(k)$ estimators as follows

$$\begin{aligned} (\widehat{\alpha}_{\mathrm{S}}, \widehat{\lambda}_{\mathrm{S}}) &= \left(\widehat{\rho}_{\mathrm{S}}(1), \overline{Y}\,[1-\widehat{\rho}_{\mathrm{S}}(1)]\right), \\ (\widehat{\alpha}_{\mathrm{G}}, \widehat{\lambda}_{\mathrm{G}}) &= \left(\widehat{\rho}_{\mathrm{G}}(1), \overline{Y}\,[1-\widehat{\rho}_{\mathrm{G}}(1)]\right). \end{aligned}$$

4.5 Numerical results

In order to compare the performances of the different robust methodologies previously discussed, we performed a simulation study for different sample sizes and for different parameter values. All simulations were carried out using the R programming language, which is freely distributed and available at http://www.r-project.org. The data set $Y_1, \ldots, Y_T$ was generated according to model (4.1) with $\{\epsilon_t\}$ being an i.i.d. Poisson sequence with mean $\lambda = 1$. The contaminated data were generated from model (4.4) with $p = 0.02$ for magnitudes $\omega = 0, 4$ and 7. The sample sizes considered were $T = 100, 200, 300$ and the values of the α parameter in the simulation study were $\alpha = 0.2$ and 0.6. For each different situation the empirical mean and mean squared error (MSE) of the estimators were numerically estimated. The values of the MSE are given between parentheses. The number of Monte Carlo replications was 5000. The results are presented in Tables 4.1 to 4.2.

Initially, the case where the series is not contaminated ($\omega = 0$) is analyzed. It can be seen that the SD estimator presents much smaller biases (in absolute values) than the other estimators, for all cases. Tables 1 and 2 show that additive outliers introduce significant positive bias in the SD estimator for λ parameter and significant negative bias in the SD estimator for α parameter.

For small sample sizes ($T = 100$) and $\omega > 0$, in general, both biases and MSE for the Spearman estimator (for α and λ) are smaller than those for the SD and YW methods. Furthermore, for $\omega = 7$, the biases for Gaussian estimators of α and λ are smaller than those for the SD, YW and Spearman methods. Another result is related to the size of α. In general, for YW, Gaussian and Spearman estimates of λ, increasing α, the bias and MSE also increase.

The empirical investigation here presented suggests that, in general, when there is no evidence of atypical observations, SD estimates give satisfactory results. The empirical study also provides evidence that special attention has to be paid when the data possibly exhibits atypical observations. In general, all the other methods (YW, Spearman and Gaussian) present pretty similar results. Significant changes in the behavior of the SD estimation method are observed when the process has outliers, especially for large ω.

Table 4.1: Empirical means and mean squared errors (in parentheses) of the estimates of the parameter α for $p = 0.02$, $\lambda = 1$ and some values of T and ω.

α	T	Estimator	$\omega = 0$	$\omega = 4$	$\omega = 7$
0.2	100	$\widehat{\alpha}_{\text{SD}}$	0.1948 (0.0186)	0.0556 (0.0554)	−0.2747 (0.3553)
		$\widehat{\alpha}_{\text{YW}}$	0.1794 (0.0108)	0.1785 (0.0113)	0.1795 (0.0109)
		$\widehat{\alpha}_{\text{S}}$	0.1718 (0.0105)	0.1780 (0.0105)	0.1923 (0.0097)
		$\widehat{\alpha}_{\text{G}}$	0.1785 (0.0104)	0.1824 (0.0104)	0.1991 (0.0096)
	200	$\widehat{\alpha}_{\text{SD}}$	0.1981 (0.0092)	0.0555 (0.0377)	−0.2729 (0.2882)
		$\widehat{\alpha}_{\text{YW}}$	0.1902 (0.0053)	0.1902 (0.0057)	0.1901 (0.0055)
		$\widehat{\alpha}_{\text{S}}$	0.1807 (0.0053)	0.1883 (0.0052)	0.2010 (0.0048)
		$\widehat{\alpha}_{\text{G}}$	0.1885 (0.0051)	0.1934 (0.0052)	0.2092 (0.0049)
	300	$\widehat{\alpha}_{\text{SD}}$	0.1973 (0.0063)	0.0572 (0.0313)	−0.2722 (0.2659)
		$\widehat{\alpha}_{\text{YW}}$	0.1921 (0.0036)	0.1947 (0.0036)	0.1933 (0.0039)
		$\widehat{\alpha}_{\text{S}}$	0.1818 (0.0036)	0.1920 (0.0033)	0.2014 (0.0033)
		$\widehat{\alpha}_{\text{G}}$	0.1900 (0.0034)	0.1976 (0.0033)	0.2107 (0.0034)
0.6	100	$\widehat{\alpha}_{\text{SD}}$	0.5976 (0.0048)	0.5443 (0.0104)	0.4247 (0.0514)
		$\widehat{\alpha}_{\text{YW}}$	0.5624 (0.0089)	0.5645 (0.0088)	0.5648 (0.0087)
		$\widehat{\alpha}_{\text{S}}$	0.5540 (0.0095)	0.5627 (0.0086)	0.5811 (0.0070)
		$\widehat{\alpha}_{\text{G}}$	0.5609 (0.0084)	0.5665 (0.0079)	0.5811 (0.0065)
	200	$\widehat{\alpha}_{\text{SD}}$	0.5992 (0.0023)	0.5463 (0.0065)	0.4210 (0.0427)
		$\widehat{\alpha}_{\text{YW}}$	0.5820 (0.0042)	0.5827 (0.0040)	0.5809 (0.0042)
		$\widehat{\alpha}_{\text{S}}$	0.5692 (0.0047)	0.5770 (0.0040)	0.5920 (0.0033)
		$\widehat{\alpha}_{\text{G}}$	0.5788 (0.0040)	0.5830 (0.0036)	0.5944 (0.0030)
	300	$\widehat{\alpha}_{\text{SD}}$	0.5999 (0.0015)	0.5453 (0.0053)	0.4221 (0.0384)
		$\widehat{\alpha}_{\text{YW}}$	0.5888 (0.0026)	0.5881 (0.0027)	0.5881 (0.0027)
		$\widehat{\alpha}_{\text{S}}$	0.5749 (0.0031)	0.5803 (0.0028)	0.5967 (0.0022)
		$\widehat{\alpha}_{\text{G}}$	0.5855 (0.0025)	0.5876 (0.0024)	0.6001 (0.0020)

Table 4.2: Empirical means and mean squared errors (in parentheses) of the estimates of the parameter λ for $p = 0.02$, $\lambda = 1$ and some values of T and ω.

α	T	Estimator	$\omega = 0$	$\omega = 4$	$\omega = 7$
0.2	100	$\widehat{\lambda}_{\text{SD}}$	1.0034 (0.0363)	1.2797 (0.1704)	1.8431 (1.1443)
		$\widehat{\lambda}_{\text{YW}}$	1.0228 (0.0248)	1.1096 (0.0444)	1.1679 (0.0696)
		$\widehat{\lambda}_{\text{S}}$	1.0322 (0.0244)	1.1097 (0.0422)	1.1482 (0.0574)
		$\widehat{\lambda}_{\text{G}}$	1.0239 (0.0241)	1.1040 (0.0414)	1.1388 (0.0547)
	200	$\widehat{\lambda}_{\text{SD}}$	1.0011 (0.0188)	1.2777 (0.1220)	1.8273 (0.8970)
		$\widehat{\lambda}_{\text{YW}}$	1.0110 (0.0128)	1.0935 (0.0247)	1.1530 (0.0437)
		$\widehat{\lambda}_{\text{S}}$	1.0229 (0.0129)	1.0960 (0.0239)	1.1366 (0.0353)
		$\widehat{\lambda}_{\text{G}}$	1.0132 (0.0125)	1.0891 (0.0229)	1.1250 (0.0324)
	300	$\widehat{\lambda}_{\text{SD}}$	1.0027 (0.0125)	1.2718 (0.1029)	1.8176 (0.8056)
		$\widehat{\lambda}_{\text{YW}}$	1.0091 (0.0083)	1.0849 (0.0176)	1.1463 (0.0348)
		$\widehat{\lambda}_{\text{S}}$	1.0220 (0.0082)	1.0885 (0.0172)	1.1344 (0.0294)
		$\widehat{\lambda}_{\text{G}}$	1.0117 (0.0080)	1.0809 (0.0161)	1.1213 (0.0259)
0.6	100	$\widehat{\lambda}_{\text{SD}}$	0.9983 (0.0330)	1.2262 (0.1207)	1.6593 (0.7073)
		$\widehat{\lambda}_{\text{YW}}$	1.0870 (0.0618)	1.1692 (0.0933)	1.2399 (0.1365)
		$\widehat{\lambda}_{\text{S}}$	1.1078 (0.0661)	1.1741 (0.0924)	1.1911 (0.1016)
		$\widehat{\lambda}_{\text{G}}$	1.0908 (0.0589)	1.1640 (0.0859)	1.1924 (0.1009)
	200	$\widehat{\lambda}_{\text{SD}}$	0.9993 (0.0167)	1.2201 (0.0819)	1.6577 (0.5708)
		$\widehat{\lambda}_{\text{YW}}$	1.0427 (0.0292)	1.1212 (0.0464)	1.1921 (0.0763)
		$\widehat{\lambda}_{\text{S}}$	1.0745 (0.0323)	1.1366 (0.0489)	1.1594 (0.0564)
		$\widehat{\lambda}_{\text{G}}$	1.0506 (0.0276)	1.1206 (0.0434)	1.1532 (0.0539)
	300	$\widehat{\lambda}_{\text{SD}}$	0.9984 (0.0108)	1.2268 (0.0737)	1.6499 (0.5096)
		$\widehat{\lambda}_{\text{YW}}$	1.0258 (0.0172)	1.1104 (0.0333)	1.1710 (0.0545)
		$\widehat{\lambda}_{\text{S}}$	1.0607 (0.0205)	1.1312 (0.0373)	1.1460 (0.0420)
		$\widehat{\lambda}_{\text{G}}$	1.0341 (0.0165)	1.1117 (0.0312)	1.1367 (0.0386)

4.6 Real data example: Counts of IP addresses

In this section, we apply the methodology considered in Section 4.4 to a real data set. The data set consists of the number of different IP addresses (≈ different users), registered within periods of 2-min length can be read, accessing the server of the Department of Statistics of the University of Würzburg on November 29th, 2005, between 10 am and 6 pm (241 observations). This series was previously studied by Weiß (2007) and Silva and Pereira (2012). The required numerical evaluations for data analysis were here implemented using the R software.

Table 4.3 displays some descriptive statistics. Some observations assume the value 0. Furthermore, we note that the ratio between the sample variance and the sample mean is 1.059, hence, the data seem to be equidispersed. Consequently, a Poisson marginal distribution seems to be appropriate. The series, sample autocorrelation and sample partial autocorrelation are displayed in Figure 4.1.

Table 4.3: Descriptive statistics

Minimum	Median	Mean	Variance	$\widehat{\rho}(1)$	Maximum
0	1	1.315	1.392	0.219	8

Analyzing Figure 4.1, 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. According to Weiß (2007) and Silva and Pereira (2012), the observation at time $t = 224$ ($Z_{224} = 8$) is a possible occurrence of an outlier with $\omega = 7$. For this application, the estimates of α and λ were computed from the original data and from the modified data (without outlier). Therefore, we replaced the outlier by defining $Z_{224} = 1$. As a consequence, the mean changes to 1.2863, the first order autocorrelation to 0.2925, but an analysis of the resulting partial autocorrelation function and of the histogram showed that a Poisson INAR(1) model is still reasonable (Weiß, 2007).

In Table 4.4 we present the estimates from the estimators discussed in Sections 4.2 and 4.4 for the two series with and without outliers. In both series, the robust methods present similar results. In contrast to the robust methods, the classics SD and YW estimators give estimates that dramatically change from original data to data without outliers, showing that the observation is a possible outlier.

Table 4.4: Estimates of the parameters for counts of IP data.

Estimator	Counts of IP data	without outlier
$(\widehat{\alpha}_{SD}, \widehat{\lambda}_{SD})$	(0.1796, 1.0792)	(0.3424, 0.8458)
$(\widehat{\alpha}_{YW}, \widehat{\lambda}_{YW})$	(0.2195, 1.0267)	(0.2925, 0.9101)
$(\widehat{\alpha}_{S}, \widehat{\lambda}_{S})$	(0.2517, 0.9843)	(0.2680, 0.9416)
$(\widehat{\alpha}_{G}, \widehat{\lambda}_{G})$	(0.2480, 0.9892)	(0.2799, 0.9263)

4.7 Conclusions

This chapter develops a robust estimation approach for the parameters that index the Poisson INAR(1) model. Furthermore, this chapter investigates the impact of additive outliers on estimating

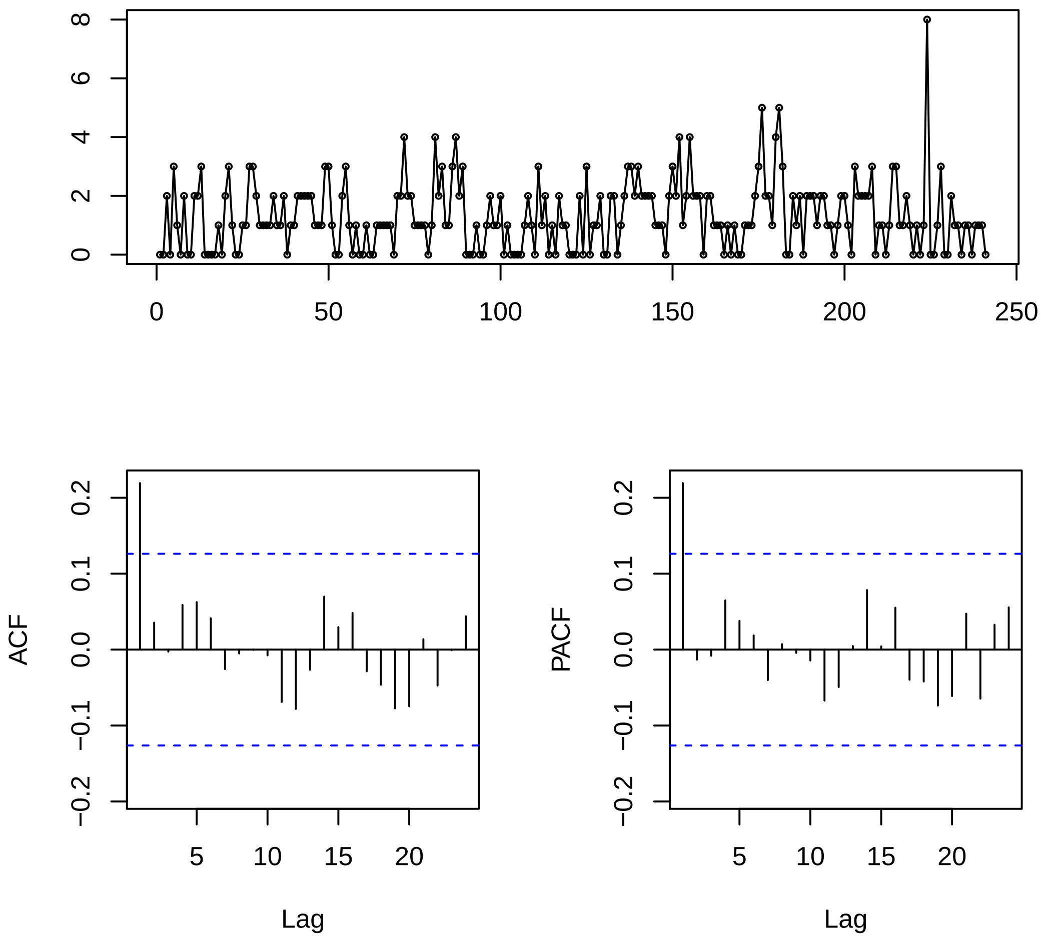


Figure 4.1: The time series plot, autocorrelation and partial autocorrelation functions for number of different IP adresses.

in the Poisson INAR(1) model. The number of different IP addresses data was analyzed as an application of the methodology studied here. The simulation results showed that the robust methods can be used as an alternative estimation for the parameters of the Poisson INAR(1) model containing additive outliers. In future research, the asymptotic properties of the robust estimators remain to be investigated.

4.8 Appendix

Proof of Proposition 4

It is easy to verify that

$$\sum_{t=2}^{T}(Z_t - Z_{t-1})^2 = \sum_{t=2}^{T}(Y_t - Y_{t-1})^2 + 2\,\omega^2 \sum_{t=2}^{T} \delta_t(\delta_t - \delta_{t-1}).$$

Thus,

$$\widehat{\lambda}_{\mathrm{SD}}^{Z} = \frac{1}{2(T-1)} \sum_{t=2}^{T}(Z_t - Z_{t-1})^2 = \widehat{\lambda}_{\mathrm{SD}}^{Y} + \frac{\omega^2 \sum_{t=2}^{T} \delta_t(\delta_t - \delta_{t-1})}{T-1}.$$

References

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

Barczy, M., Ispány, M., Pap, G., Scotto, M., Silva, M.E. (2010). Innovational outliers in INAR(1) models. *Communications in Statistics - Theory and Methods*, **39**, 3343–3362.

Barczy, M., Ispány, M., Pap, G., Scotto, M., Silva, M.E. (2012). Additive outliers in INAR(1) models. *Statistical Papers*, **53**, 935–949.

Blomqvist, N. (1950). On a measure of dependence between two random variables. *The Annals of Mathematical Statistics*, **21**, 593–600.

Bourguignon, M. and Vasconcellos, K.L.P. (2014). Improved estimation for Poisson INAR(1) models. *Journal of Statistical Computation and Simulation*. accept. DOI: 10.1080/00949655.2014.930862.

Boudt, K., Cornelissen, J. and Croux, C. (2012). The Gaussian rank correlation estimator: robustness properties. *Stat. Comput.* **22**, 471–483.

Billingsley, P. *Probability and Measure*, John Wiley & Sons, New York, 1979.

Brännäs, K. (1994). Estimation and testing in integer-valued AR(1) models. *Umeå Economic Studies*, **335**. http://www.usbe.umu.se/digitalAssets/39/39107_ues335.pdf Accessed 04 September 2012.

Brännäs, K., Hellström, J. and Nordström, J. (2002) . A new approach to modelling and forecasting monthly guest nights in hotels. *Int. J. Forecasting*, **18**, 19–30.

Denby, L. and Martin, R.D. (1979). Robust estimation of the first order autoregressive parameter. *Journal of the American Statistical Association*, **74**, 140–146.

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

Fajardo, F; Reisen, V.A. and Cribari-Neto, F. (2009) Robust estimation in long-memory processes under additive outliers. *Journal of Statistical Planning Inference*, **139**, 2511–2525.

Franke, J. and Seligmann, T. (1993). Conditional maximum-likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In: Subba Rao T, editor. Developments in time series. London (UK): Chapman & Hall; p. 310–330.

Fox, A.J. (1972). Outliers in time series. *Journal of the Royal Statistical Society B*, **34**, 350–363.

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

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

Hájek, J. and Sidak, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

Hellström, J. (2001). Unit root testing in integer-valued AR(1) models. *Economics Letters*, **70**, 9–14.

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

Kachour, M., and Yao, J. F. (2009). First-order rounded integer-valued autoregressive (RINAR(1)) process. *Journal of Time Series Analysis*, **30**, 417–448.

Martin, R.D. and Yohai, V.J. (1985). Robustness in time series and estimating ARMA models. In*Handbook of Statistics*, Vol. 5, edited by E. J. Hannan, et al., 119155, *Science Publishers B.V.*

Park, Y. and Oh, C.W. (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals. *Statistical Papers*, **38**, 287–302.

Silva, M.E. and Pereira, I. (2012). Detection of additive outliers in Poisson Integer-valued AutoRegressive time series. Preprint: http://arxiv.org/pdf/1204.6516.pdf

Sarnaglia, A.J.Q.; Reisen, V.A. and Lévy-Leduc, C. (2010). Robust estimation of periodic autoregressive processes in the presence of additive outliers. *Journal of Multivariate Analysis*, **101**, 2168–2183.

Spearman, C. (1904). General intelligence objectively determined and measured. *Am. J. Psychol.*, **15**, 201–293.

Steutel, F.W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, **7**, 893–899.

Weiß, C.H. (2007). Controlling correlated processes of Poisson counts. *Qual. Reliab. Engng. Int.*, **23**, 741–754.

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

Weiß, C.H. (2011). Simultaneous confidence regions for the parameters of a Poisson INAR(1) model. *Statistical Methodology*, **8**, 517–527.

Weiß, C.H. (2012). Process capability analysis for serially dependent processes of Poisson counts. *Journal of Statistical Computation and Simulation*, **82**, 383–404.

Zeger, S.L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, **44**, 1019–1031.

CHAPTER 5

A new skew integer valued time series process

Resumo

Neste capítulo, introduzimos o processo autorregressivo de primeira ordem para valores inteiros (positivos e negativos) com distribuição marginal geométrica–Poisson. Várias propriedades do processo são estabelecidas. Os parâmetros desconhecidos do processo são estimados utilizando o método de Yule-Walker e suas propriedades assintóticas são consideradas. Alguns resultados numéricos dos estimadores são apresentados com uma breve discussão. A flexibilidade do novo modelo é ilustrada com uma aplicação a um conjunto de dados reais.

Palavras-chave: Estimador de Yule-Walker. Operador thinning; Processos INAR(1). Séries temporais de valores inteiros.

Abstract

In this chapter, we introduce a stationary first-order integer-valued autoregressive process with geometric-Poisson marginals. The new process allows negative values for the series. Also, the new process has as a particular case the NGINAR(1) process. Several properties of the process are established. The unknown parameters of the model are estimated using the Yule-Walker method and the asymptotic properties of the estimator are considered. Some numerical results of the estimators are presented with a brief discussion. Possible application of the process is discussed through a real data example.

Keywords: INAR(1) process. Integer-valued time series. Thinning operator; Yule-Walker.

5.1 Introduction

Integer-valued times series with support in $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ are very frequent in practice. The need for modelling and analysing counting data (negative and positive) occurs in many fields of real life such as medicine (Karlis and Ntzoufras, 2006), image analysis (Hwang et al., 2007), sports applications (Karlis and Ntzoufras, 2009), financial applications (Alzaid and Omair, 2010) and so on.

Models for time series defined on the set $\mathbb{Z}$ of integers have been discussed by various researchers. Kim and Park (2008) introduced an integer-valued autoregressive process of order p with signed binomial thinning operator (INARS(p)). Zhang et al. (2010) introduced pth-order integer valued autoregressive processes with a signed generalized power series thinning operator. Kachour and Truquet (2011) introduced a more general class, based on a modified version of the generalized thinning operator, also called the signed thinning operator. Kachour and Yao (2009) proposed the RINAR model, based on the rounding operator. Some of these models arise as the difference between two discrete distributions. For example, Freeland (2010) defined the true integer-valued autoregressive process of order one (TINAR(1)) as the difference of two Poisson INAR(1) processes which requires observing the two processes. Recently, Barreto-Souza and Bourguignon (2014) introduced a stationary AR(1) process on $\mathbb{Z}$ with skew discrete Laplace marginals (which are distributed as a difference between two geometric random variables); this model is named skew INAR(1) process on $\mathbb{Z}$ (SINARZ(1)).

In a similar manner, we introduce a stationary first-order integer-valued autoregressive process with geometric–Poisson marginals (which are distributed as a difference between geometric and Poisson random variables), named new skew INAR(1) process (NSINAR(1)). Additionally, we will provide a comprehensive account of the mathematical properties of the proposed new process. The motivation for such a process arises from its potential in modelling and analyzing integer valued time series for which the non-negative part presents greater overdispersion than the negative part. In this context, it is plausible to consider the difference between two distinct processes, one with greater overdispersion for the positive part (which we wil model as geometric) and another with lower variability for the negative part (which we will model as Poisson).

The rest of the chapter unfolds as follows. In Section 5.2, the NSINAR(1) process is introduced. In Section 5.3, some of its properties are outlined. Estimation methods for the model parameters are proposed, while numerical results from Monte Carlo simulation experiments are presented and discussed in Section 5.4. An application to a real data set is addressed in Section 5.5. Finally, in Section 5.6, we offer some concluding remarks.

5.2 NSINAR(1) process

Let $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$ denote the set of non-negative integers, integers and real numbers, respectively. All random variables will be defined on the same probability space. In this section, we introduce a stationary first-order integer-valued autoregressive process with geometric–Poisson marginals, named new skew INAR(1) process. With this aim, we first review the NGINAR(1) process by Ristić et al. (2009) and the Poisson INAR(1) process by Al-Osh and Alzaid (1987).

The NGINAR(1) process is defined such that

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

where " $*$ " denotes the negative binomial thinning operator (Ristić et al., 2009). This operator is defined by $\alpha * X \overset{d}{=} \sum_{i=1}^{X} W_i$, where the symbol "$\overset{d}{=}$" means "has the same distribution as" and $\{W_i\}_{i=1}^{\infty}$ is a sequence of independent and identically distributed (i.i.d.) random variables, independent of X, following a geometric distribution with mean $\alpha \in [0, 1)$. Also, $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables independent of $\{W_i\}_{i=1}^{\infty}$, with ϵ_t and X_{t-l} being independent for all $l \geq 1$ and distributed such that $\{X_t\}_{t\in\mathbb{Z}}$ is a stationary process having geometric marginals with probability function assuming the form $\Pr(X_t = x) = \mu^x/(1+\mu)^{x+1}$, for all non-negative integer x, where $\mu > 0$.

Ristić et al. (2009) proved that the probability mass function of ϵ_t is given by

$$\Pr(\epsilon_t = l) = \left(1 - \frac{\alpha\mu}{\mu - \alpha}\right) \frac{\mu^l}{(1+\mu)^{l+1}} + \frac{\alpha\mu}{\mu - \alpha} \frac{\alpha^l}{(1+\alpha)^{l+1}}, \quad l \in \mathbb{N}, \tag{5.1}$$

with the condition $\alpha \leq \mu/(1+\mu)$ being necessary to guaranteee that all probabilities in (5.1) are nonnegative. The distribution of the random variable ϵ_t is therefore a mixture of two independent geometric distributions with means μ and α. For more details about the NGINAR(1) process, see Ristić et al. (2009), Bakouch (2010), Nastić and Ristić (2012) and Nastić et al. (2012).

The Poisson INAR(1) process introduced by Al-Osh and Alzaid (1987) is defined on the discrete support $\mathbb{N}$ of nonnegative integers by means of the difference equation

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{5.2}$$

where the operator " $\circ$ " is the binomial thinning operator (Steutel and van Harn, 1979), which is defined by $\alpha \circ Y \stackrel{d}{=} \sum_{i=1}^{Y} U_i$, $\{U_i\}_{i=1}^{\infty}$ being i.i.d. random variables, independent of Y, with $\Pr(U_i = 1) = 1 - \Pr(U_i = 0) = \alpha$. Also, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. non-negative integer valued random variables which we assume to be Poisson distributed with mean $\lambda(1-\alpha)$, ε_t and Y_{t-i} being independent for all $i \geq 1$ and all t. Thus, $\{Y_t\}_{t\in\mathbb{Z}}$ is a stationary process having Poisson marginals with mean λ. The process $\{Y_t\}_{t\in\mathbb{Z}}$ satisfying (5.2) is second-order stationary if $0 \leq \alpha < 1$ (Du and Li, 1991). For more details about the Poisson INAR(1) process, see Al-Osh and Alzaid (1987), Freeland and McCabe (2004), Silva and Oliveira (2004) and Weiß (2008).

Now, let us consider two random variables X and Y in $\mathbb{N}$ and their difference $Z = X - Y$. The probability function of the difference Z has $\mathbb{Z}$ as support. If X and Y follow independently Geometric(p) and Poisson(λ) distributions with means $\mu = p/(1-p) > 0$ and $\lambda > 0$, respectively, then the random variable $Z = X - Y$ has probability mass function given by

$$\Pr(k;\mu,\lambda) \equiv \Pr(Z = k) = \frac{\mu^k e^{-\lambda/(\mu+1)}}{(1+\mu)^{k+1}} \begin{cases} 1, & k \geq 0, \\ \frac{\gamma(-k,\lambda\,\mu/(\mu+1))}{\Gamma(-k)}, & k < 0, \end{cases} \tag{5.3}$$

where $\gamma(a,z) = \frac{1}{\Gamma(a)} \int_0^z t^{a-1} e^{-t} dt$ denotes the incomplete gamma function and $\Gamma(\cdot)$ is the gamma function.

We will denote this distribution as GP(μ,λ). Ong et al. (2008) studied the distribution arising from the difference of two discrete random variables belonging to the Panjer family of distributions. They discussed some distributional properties and computation of probabilities. As a special case they defined the difference of the negative binomial and Poisson distributions. The geometric–Poisson distribution is a particular case of this difference.

Suppose that the random variable Z follows a GP(μ,λ) distribution. The moment generating function (mgf) of Z, denoted by $\Phi_Z(s) = E[\exp(s\,Z)]$, is given by

$$\Phi_Z(s) = \frac{\exp[\lambda(e^{-s} - 1)]}{[1 + \mu(1 - e^s)]}, \quad s \in \mathbb{R}. \tag{5.4}$$

It is immediate from (5.4) that the moments of Z are given by

$$E(Z^n) = \frac{\partial^n}{\partial s^n} \frac{\exp[\lambda(e^{-s} - 1)]}{[1 + \mu(1 - e^s)]}\Bigg|_{s=0}.$$

In particular, the first four moments of Z are

$$E(Z) = \mu - \lambda, \quad E(Z^2) = 2\mu^2 + \mu + \lambda^2 + \lambda - 2\mu\lambda,$$
$$E(Z^3) = 6\mu^3 - 6\lambda\mu^2 + 6\mu^2 + 3\mu\lambda^2 - 3\lambda^2 - \lambda^3 - \lambda + \mu,$$

and

$$\begin{aligned}E(Z^4) = {} & \lambda + 7\lambda^2 - 2\lambda\mu + 6\lambda^2\mu + \lambda^4 - 4\lambda^3\mu + 12\lambda^2\mu^2 \\ & - 24\lambda\mu^3 + 24\mu^4 + 36\mu^3 + 14\mu^2 + \mu.\end{aligned}$$

Further calculations show that the variance, skewness and kurtosis of Z are given by

$$\mathrm{Var}(Z) = \mu(1+\mu) + \lambda, \quad \gamma_1 = \frac{[\mu(1+\mu)(1+2\mu) - \lambda]}{[\mu(1+\mu)+\lambda]^{3/2}} \quad \text{and} \quad \gamma_2 = 3 + \frac{[\mu(1+\mu)(6\mu^2+6\mu+1)+\lambda]}{[\mu(1+\mu)+\lambda]^2},$$

respectively. Thus,

$$\frac{\gamma_2 - 3}{\gamma_1^2} = \frac{[\mu(1+\mu)(6\mu^2+6\mu+1)+\lambda][\mu(1+\mu)+\lambda]}{[\mu(1+\mu)(1+2\mu)-\lambda]^2}.$$

Clearly, $\mathrm{Var}(Z) > |E(Z)|$ and $\gamma_2 - 3 > \gamma_1^2$. It is also interesting to observe that $\gamma_1 > 0$ when $\lambda = \mu$.

Let Z be a random variable with $GP(\mu, \lambda)$ distribution. Thus, $Z \overset{d}{=} X - Y$, where X and Y are two independent random variables with Geometric$(\mu/(1+\mu))$ and Poisson(λ) distributions, respectively. We define our thinning operator " $\star$ " by

$$(\alpha \star Z)|Z \overset{d}{=} (\alpha * X - \alpha \circ Y)|(X - Y),$$

where the counting series in $\alpha * X$ and $\alpha \circ Y$ are mutually independent random variables, the counting series in $\alpha * X$ being geometrically distributed with mean α and those in $\alpha \circ Y$ being Bernoulli distributed with mean α.

Now, let X and Y be two independent random variables with Geometric$(\mu/(1+\mu))$ and Poisson(λ) distributions, respectively. It is immediate, from the above definition, that

$$\alpha \star Z \overset{d}{=} \alpha * X - \alpha \circ Y.$$

The basic properties of the thinning operator " $\star$ " are stated in Proposition 5.

Proposition 5. *Let Z be a random variable with $GP(\mu, \lambda)$ distribution.*
(i) $0 \star Z = 0$.
(ii) $1 \star Z$ *is* not *distributed as Z.*
(iii) $E(\alpha \star Z) = \alpha(\mu - \lambda)$.
(iv) $\mathrm{Var}(\alpha \star Z) = \alpha\,\mu(1 + 2\,\alpha + \alpha\,\mu) + \alpha\,\lambda$.
If X and Y are independent, X geometrically distributed with mean μ, Y Poisson with mean λ,
(v) $E(\alpha * X - \alpha \circ Y | X = x, Y = y) = \alpha(x - y)$.
(vi) $\mathrm{Var}(\alpha * X - \alpha \circ Y | X = x, Y = y) = \alpha(1+\alpha)x + \alpha(1-\alpha)y$.
(vii) $E(\alpha * X - \alpha \circ Y | X - Y = z) = \alpha z$.

Under this operator, the new skew INAR(1) process can be defined.

Definition 5.2.1. *(NSINAR(1) process) Let $\{Z_t\}_{t\in\mathbb{Z}}$ be a sequence of random variables following a common geometric–Poisson distribution with parameters μ and λ, and suppose that ξ_t, for $t \in \mathbb{Z}$, has the distribution of $\epsilon_t - \varepsilon_t$, where the sequences $\{\epsilon_t\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are independent, $\{\epsilon_t\}_{t\in\mathbb{Z}}$ are i.i.d. random variables with common probability function given in (5.1) and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are i.i.d. random variables with common Poisson$(\lambda(1-\alpha))$ distribution. We also suppose that ξ_t and Z_{t-l} are independent for all $l \geq 1$. Under this, we define completely our NSINAR(1) process $\{Z_t\}_{t\in\mathbb{Z}}$ by*

$$Z_t = \alpha \star Z_{t-1} + \xi_t, \quad t \in \mathbb{Z}. \tag{5.5}$$

From the results of Du and Li (1991) and Ristić et al. (2009) we have that $0 \leq \alpha \leq \mu/(1+\mu)$ and $\alpha \in (\mu/(1+\mu), 1]$ are the conditions of stationarity and non-stationarity of the process $\{Z_t\}_{t\in\mathbb{Z}}$, respectively. Also, $\alpha = 0$ and $\alpha > 0$, respectively, imply the independence and dependence of the observations of $\{Z_t\}_{t\in\mathbb{Z}}$. Here, we restrict our study to the stationary case.

Proposition 6. *The distribution $\alpha \star Z_{t-1}|Z_{t-1}$ is given by*

$$\Pr(\alpha \star Z_{t-1} = j|Z_{t-1} = k) = \frac{e^{-\frac{\lambda\mu}{1+\mu}}\alpha^j}{(1+\alpha)^{k+j}} \sum_{l=0}^{\infty}\sum_{m=0}^{l} \binom{j+l+k-1}{j}\binom{l}{m} \frac{\alpha^m(1-\alpha)^{l-m}}{(1+\alpha)^l} \frac{[\lambda\,\mu/(1+\mu)]^l}{l!},$$

for $k, j \geq 0$.

Remark 1. *If $\alpha = 0$, then $Z_t = \xi_t$ for all t and $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables having the $GP(\mu, \lambda)$ distribution.*

Remark 2. *If $\lambda = 0$, $\{Z_t\}_{t\in\mathbb{Z}}$ becomes the NGINAR(1) process proposed by Ristić et al. (2009).*

Figure 5.1 displays 100 simulated values of the NSINAR(1) process for different values of the parameters α, μ and λ.

5.3 Properties of the process

In this section we will consider some properties of the NSINAR(1) process, such as the conditional moments, the autocorrelation structure and innovation structure.

Using the definition of ϵ_t and equation (5.1), we obtain that the probability mass function of ξ_t can be expressed by

$$\Pr(\xi_t = k) = \left(1 - \frac{\alpha\mu}{\mu-\alpha}\right)\Pr(k;\mu,\lambda(1-\alpha)) + \frac{\alpha\mu}{\mu-\alpha}\Pr(k;\alpha,\lambda(1-\alpha)),$$

for $k \in \mathbb{Z}$, where $\Pr(k;\cdot,\cdot)$ is defined in (5.3). Note that the condition $0 \leq \alpha \leq \mu/(1+\mu)$ assures that the probabilities of the preceding mixture distributions are well defined.

That is, the random variable ξ_t is distributed as a mixture of geometric–Poisson random variables. Using this, it is straighforward to obtain that the mgf of ξ_t, denoted by $\Phi_\xi(s) = \mathrm{E}[\exp(s\,\xi_t)]$, is given by

$$\Phi_\xi(s) = \frac{[1+\alpha(1+\mu)(1-e^s)]\exp[\lambda(1-\alpha)(e^{-s}-1)]}{[1+\mu(1-e^s)][1+\alpha(1-e^s)]},$$

for $s \in \mathbb{R}$. We have that the two first cumulants of ξ_t are given by

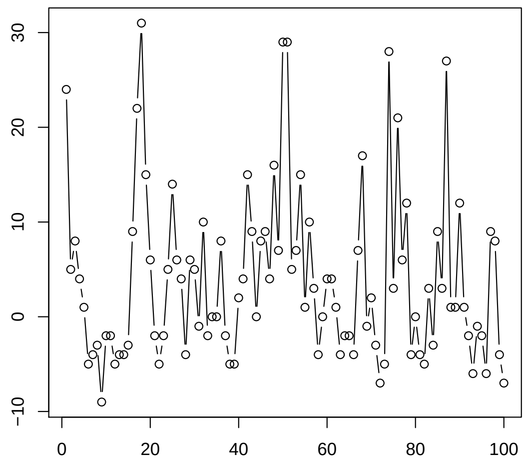
$$\begin{aligned} \mathrm{E}(\xi_t) &= \mu_\xi = (1-\alpha)(\mu-\lambda), \\ \mathrm{Var}(\xi_t) &= \sigma_\xi^2 = (1+\alpha)\mu[(1-\alpha)(1+\mu)-\alpha] + \lambda(1-\alpha). \end{aligned}$$

Remark 3. *If $\alpha = \mu/(1+\mu)$, then ξ_t has a geometric–Poisson distribution with parameters $\mu/(1+\mu)$ and $\lambda/(1+\mu)$.*

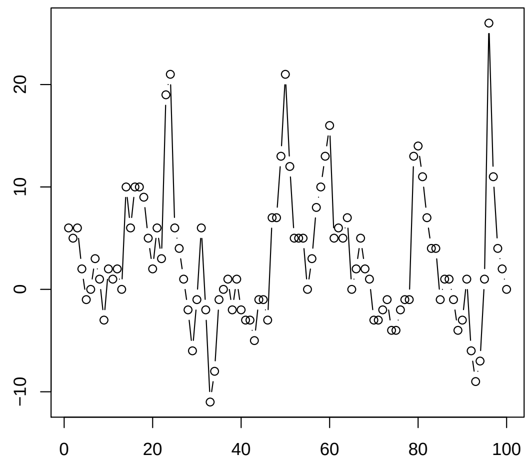
Proposition 7. *The NSINAR(1) process $\{Z_t\}_{t\in\mathbb{Z}}$ given by (5.5) is a strict stationary and ergodic process.*

The conditional moment generating function (cmgf) of Z_t given Z_{t-1} is directly derived using (5.5) and basic relations of the conditional expectation. So, for $z \geq 0$, we get

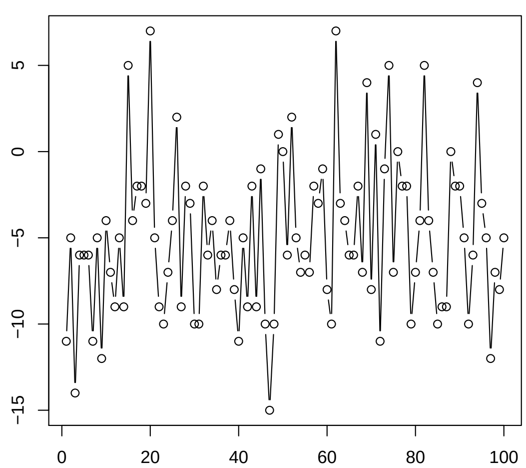
$$\mathrm{E}(e^{s\,Z_t}|Z_{t-1} = z) = \Phi_\xi(s)[1+\alpha(1-e^s)]^{-z}\exp\left\{\frac{[\lambda\,\mu/(1+\mu)][(1-\alpha)+\alpha\,e^{-s}]}{1+\alpha(1-e^s)} - \frac{\lambda\,\mu}{1+\mu}\right\}.$$



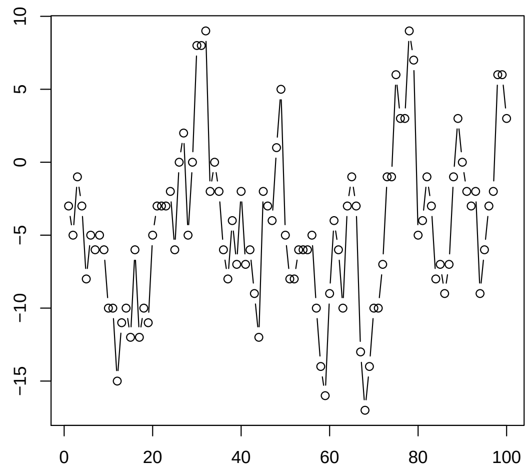
(a) $\alpha = 0.3$, $\mu = 8$ and $\lambda = 4$.



(b) $\alpha = 0.7$, $\mu = 8$ and $\lambda = 4$.



(c) $\alpha = 0.3$, $\mu = 4$ and $\lambda = 8$.



(d) $\alpha = 0.7$, $\mu = 4$ and $\lambda = 8$.

Figure 5.1: Sample paths of NSTINAR(1) process for some values of the parameters.

Now, based on the cmgf properties, the conditional expectation and conditional variance of the NSINAR(1) process for given Z_{t-1} are obtained, respectively, as

$$\begin{aligned} E(Z_t|Z_{t-1}) &= \alpha Z_{t-1} + \mu_\xi, \\ \mathrm{Var}(Z_t|Z_{t-1}) &= \alpha(1+\alpha)|Z_{t-1}| + \frac{2\,\alpha\,\lambda\,\mu}{1+\mu} + \sigma_\xi^2. \end{aligned}$$

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

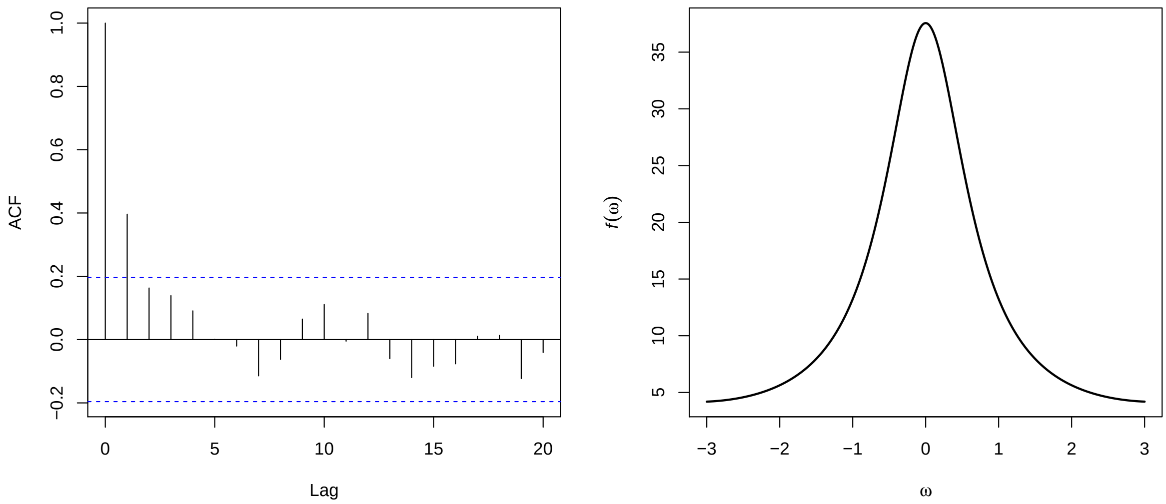
$$\mathrm{Corr}(Z_t, Z_{t-h}) = \rho(h) = \alpha^h, \quad h \geq 0, \tag{5.6}$$

for $0 \leq \alpha \leq \mu/(1+\mu)$. That is, the autocorrelation function decays exponentially as $h \to \infty$.

The spectral density function $f(\omega)$ of any order contains important information about the properties of the process. The spectral density of the NSINAR(1) process is

$$f(\omega) = \frac{\mu(1+\mu)(1-\alpha^2) + \lambda(1-\alpha^2)}{2\pi[1+\alpha^2 - 2\alpha\cos(\omega)]}, \quad \omega \in (-\pi, \pi].$$

Figure 5.2 displays a sample autocorrelation based on 100 simulated values and the spectral density function for $\alpha = 0.5, \mu = 8$ and $\lambda = 4$.



(a) Sample ACF of NSINAR(1) process.

(b) The spectral density of NSINAR(1) process.

Figure 5.2: Sample autocorrelations and spectral density of the NSINAR(1) process.

5.4 Estimation and inference of the unknown parameters

This section is concerned with the estimation of the three parameters of interest. Since the likelihood for this process is cumbersome to work with, we will use Yule-Walker to estimate the model parameters.

From a sample $Z_1, \ldots, Z_T$, the Yule-Walker (YW) estimator of α, based upon the fact that $\rho(h) = \alpha^h$, as in (5.6), is given by

$$\widehat{\alpha} = \frac{\sum_{t=1}^{T-1}(Z_t - \overline{Z})(Z_{t+1} - \overline{Z})}{\sum_{t=1}^{T}(Z_t - \overline{Z})^2},$$

where $\overline{Z} = (1/T)\sum_{t=1}^{T} Z_t$ is the sample mean. We have that the first two cumulants of Z_t are given by $\mathrm{E}(Z_t) = \mu - \lambda$ and $\mathrm{Var}(Z_t) = \mu(1+\mu) + \lambda$. Using this, the moment estimators of μ and λ are defined as

$$\widehat{\mu} = \sqrt{\overline{Z} + \widehat{\gamma}(0) + 1} - 1 \ \text{ and } \ \widehat{\lambda} = \sqrt{\overline{Z} + \widehat{\gamma}(0) + 1} - \overline{Z} - 1 = \widehat{\mu} - \overline{Z},$$

respectively, where $\widehat{\gamma}(h) = (1/T)\sum_{t=1}^{T-h}(Z_t - \overline{Z})(Z_{t-h} - \overline{Z})$ is the autocovariance function. Note that the moment estimate for μ is well defined if $\overline{Z} + \widehat{\gamma}(0) > 0$. On the other hand, both moment estimates are well defined for $\overline{Z} > 0$, if $\widehat{\gamma}(0) > \overline{Z}(\overline{Z}+1)$, and for $\overline{Z} < 0$, if $\overline{Z} + \widehat{\gamma}(0) > 0$. In simulated samples, cases like this usually happen when μ is very small compared to λ. Figure 5.3 displays the plots of these regions for the estimates.

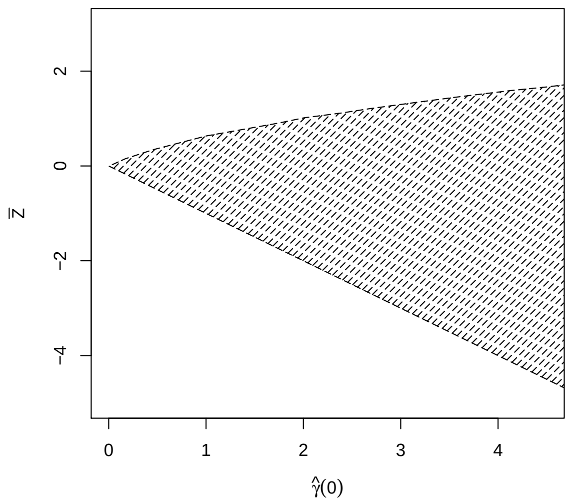


Figure 5.3: The plots for the defined regions.

Now, we can derive the asymptotic properties of the YW estimators. They follow from the following proposition, whose proof can be found in the Appendix.

Proposition 8. *The estimator* $(\widehat{\alpha}, \widehat{\mu}, \widehat{\lambda})$ *is strongly consistent and has an asymptotic normal distribution.*

Another idea that is frequently considered in INAR modelling is to estimate the unknown parameters by minimizing

$$\sum_{t=2}^{T} |Z_t - \mathrm{E}[Z_t|Z_{t-1}]|^p,$$

for a given $p \geq 1$. For example, the case $p = 2$ will give us conditional least squares estimation. However, in our case, $\mathrm{E}[Z_t|Z_{t-1}] = \alpha\, Z_{t-1} + (1-\alpha)(\mu - \lambda)$, which depends on μ and λ only through $\mu - \lambda$. Therefore, we cannot define estimators of μ and λ with this technique, in our case. For this reason, this kind of estimation was not were considered.

Next, a small Monte Carlo simulation experiment will be conducted to evaluate the estimation of the NSINAR(1) process parameters. The simulation was performed using the R programming language; see http://www.r-project.org. The number of Monte Carlo replications was $R = 5000$. The sample sizes considered are $T = 100, 200, 400, 800$ and the values of the parameters are the combinations $\{(\mu,\lambda) = (8,4), (\mu,\lambda) = (4,8)\} \times \{\alpha = 0.15, \alpha = 0.30, \alpha = 0.45, \alpha = 0.60, \alpha = 0.75\}$.

Tables 5.1 and 5.2 present the empirical mean and mean squared error of the the estimates of the parameters of the NSINAR(1) process. From those tables, notice that the biases and mean square errors decrease as the size of the sample increases, evidencing that the estimators are asymptotically unbiased.

Table 5.1: Empirical means and mean squared errors (in parentheses) of the estimates of the parameters for $(\mu, \lambda) = (8, 4)$ and some values of α and T.

T	α	$\widehat{\alpha}$	$\widehat{\mu}$	$\widehat{\lambda}$
100	0.15	0.1332 (0.0097)	7.9002 (1.3816)	3.9156 (0.7485)
	0.30	0.2795 (0.0097)	7.8661 (1.6850)	3.8730 (0.8149)
	0.45	0.4222 (0.0094)	7.8425 (2.0949)	3.8139 (1.0496)
	0.60	0.5615 (0.0092)	7.6840 (2.9914)	3.6977 (1.4583)
	0.75	0.7038 (0.0086)	7.5395 (5.0219)	3.5091 (2.3491)
200	0.15	0.1409 (0.0048)	7.9463 (0.7364)	3.9642 (0.3708)
	0.30	0.2878 (0.0050)	7.9207 (0.8633)	3.9209 (0.4209)
	0.45	0.4360 (0.0045)	7.9192 (1.1075)	3.9229 (0.5629)
	0.60	0.5821 (0.0043)	7.8800 (1.5755)	3.8594 (0.7425)
	0.75	0.7244 (0.0040)	7.6911 (2.6732)	3.7210 (1.2557)
400	0.15	0.1453 (0.0025)	7.9705 (0.3533)	3.9790 (0.1845)
	0.30	0.2941 (0.0024)	7.9724 (0.4241)	3.9723 (0.2118)
	0.45	0.4419 (0.0023)	7.9458 (0.5512)	3.9426 (0.2697)
	0.60	0.5909 (0.0022)	7.9228 (0.8386)	3.9164 (0.4027)
	0.75	0.7374 (0.0019)	7.8387 (1.4162)	3.8559 (0.6256)
800	0.15	0.1473 (0.0013)	7.9891 (0.1910)	3.9873 (0.0986)
	0.30	0.2967 (0.0013)	7.9750 (0.2238)	3.9806 (0.1116)
	0.45	0.4464 (0.0012)	7.9736 (0.2756)	3.9658 (0.1372)
	0.60	0.5956 (0.0010)	7.9581 (0.4003)	3.9549 (0.1927)
	0.75	0.7435 (0.0009)	7.9512 (0.7465)	3.9162 (0.3271)

Table 5.2: Empirical means and mean squared errors (in parentheses) of the estimates of the parameters for $(\mu, \lambda) = (4, 8)$ and some values of α and T.

T	α	$\widehat{\alpha}$	$\widehat{\mu}$	$\widehat{\lambda}$
100	0.15	0.1327 (0.0098)	3.9273 (0.4559)	7.9270 (0.3355)
	0.30	0.2758 (0.0100)	3.9364 (0.5592)	7.9157 (0.4172)
	0.45	0.4225 (0.0098)	3.8875 (0.6874)	7.8787 (0.5400)
	0.60	0.5638 (0.0092)	3.8163 (1.0493)	7.8069 (0.7827)
	0.75	0.6995 (0.0094)	3.6673 (1.7430)	7.6685 (1.2797)
200	0.15	0.1438 (0.0051)	3.9814 (0.2281)	7.9753 (0.1738)
	0.30	0.2880 (0.0049)	3.9653 (0.2793)	7.9706 (0.2171)
	0.45	0.4347 (0.0048)	3.9390 (0.3479)	7.9343 (0.2772)
	0.60	0.5806 (0.0045)	3.9016 (0.5462)	7.9095 (0.3942)
	0.75	0.7241 (0.0040)	3.8203 (0.9354)	7.8166 (0.6349)
400	0.15	0.1463 (0.0025)	3.9885 (0.1167)	7.9908 (0.0867)
	0.30	0.2942 (0.0025)	3.9875 (0.1409)	7.9805 (0.1069)
	0.45	0.4443 (0.0024)	3.9798 (0.1822)	7.9746 (0.1364)
	0.60	0.5908 (0.0021)	3.9509 (0.2612)	7.9563 (0.2027)
	0.75	0.7359 (0.0019)	3.8994 (0.4835)	7.9043 (0.3407)
800	0.15	0.1480 (0.0012)	3.9928 (0.0571)	7.9941 (0.0445)
	0.30	0.2977 (0.0012)	3.9884 (0.0682)	7.9900 (0.0533)
	0.45	0.4465 (0.0012)	3.9853 (0.0914)	7.9848 (0.0700)
	0.60	0.5949 (0.0011)	3.9680 (0.1435)	7.9683 (0.1013)
	0.75	0.7426 (0.0009)	3.9473 (0.2473)	7.9498 (0.1778)

5.5 Application to the Saudi stock market data

In this section, we present an application of the NSINAR(1) model to the Saudi stock market. We consider the daily difference between the close and open prices in 2012 of the Saudi Telecom. Since these differences belong to $\mathbb{Z}/10$, we work with the rescaled time series (close price $-$ open price)$\times$ 10, which we call number of ticks. The number of ticks belongs to $\mathbb{Z}$ and is our time series of interest here in this section. The data consist of 251 observations. This data set was recently used in Barreto-Souza and Bourguignon (2014). The data has been downloaded from the Saudi Stock Exchange (`http://www.tadawul.com.sa`).

Table 5.3 provides some descriptive measures for the number of ticks, including central tendency statistics, the variance, skewness and kurtosis, among others. We see that the data set assumes positive and negative integer values.

Table 5.3: Descriptive statistics.

Minimum	Median	Mean	Variance	Skewness	Kurtosis	$\widehat{\rho}(1)$	Maximum
-18	0	0.5259	18.066	0.9337	8.1169	0.1795	22

The time series data and their sample autocorrelation and partial autocorrelation are displayed in Figure 5.4. Analyzing Figure 5.4, 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 following ones: the SINARZ(1) process introduced by Barreto-Souza and Bourguignon (2014) and the TINAR(1) process introduced by Freeland (2010) (asymmetric version).

In Table 5.4 we present the estimates of the parameters and three goodness-of-fit statistics are also presented: the RMS (root mean-squared error), MA (mean absolute error) and MdA (median absolute error). Those statistics are defined as follows. For $t = 2, \ldots, T$, consider the estimated expected value of the observation at the previous time, $\widehat{\mathrm{E}}[Z_t|Z_{t-1}] = \widehat{\alpha}\, Z_{t-1} + (1-\widehat{\alpha})(\widehat{\mu} - \widehat{\lambda})$ for NSINAR(1). The RMS is obtained as the square root of the average value of $[Z_t - \widehat{\mathrm{E}}[Z_t|Z_{t-1}]]^2$, the MA is obtained as the average value of $|Z_t - \widehat{\mathrm{E}}[Z_t|Z_{t-1}]|$ and the MdA is obtained as the median of the values of $|Z_t - \widehat{\mathrm{E}}[Z_t|Z_{t-1}]|$. In general it is expected from the best model to fit the data to present the smaller values for these statistics. From this table, we observe that the three models are competitive, with the first being marginally better. Hence, our proposed model seems to be a good alternative. The sample autocorrelations of the residuals obtained from NSINAR(1) model are shown in Figure 5.5. From this figure, we conclude that the proposed fitted the data well.

5.6 Concluding remarks

In this chapter, we introduce a stationary first-order integer-valued autoregressive process with geometric–Poisson marginals. The new model has several advantages: possible negative values for time series and simple innovation structure. The main properties of the model are derived. Then we considered the problem of parameter estimation and derived the asymptotic normality distribution of the Yule-Walker estimators. An example of application to a real data set illustrates the importance and potentiality of the new model. We hope this new process may attract wider applications in count time series analysis.

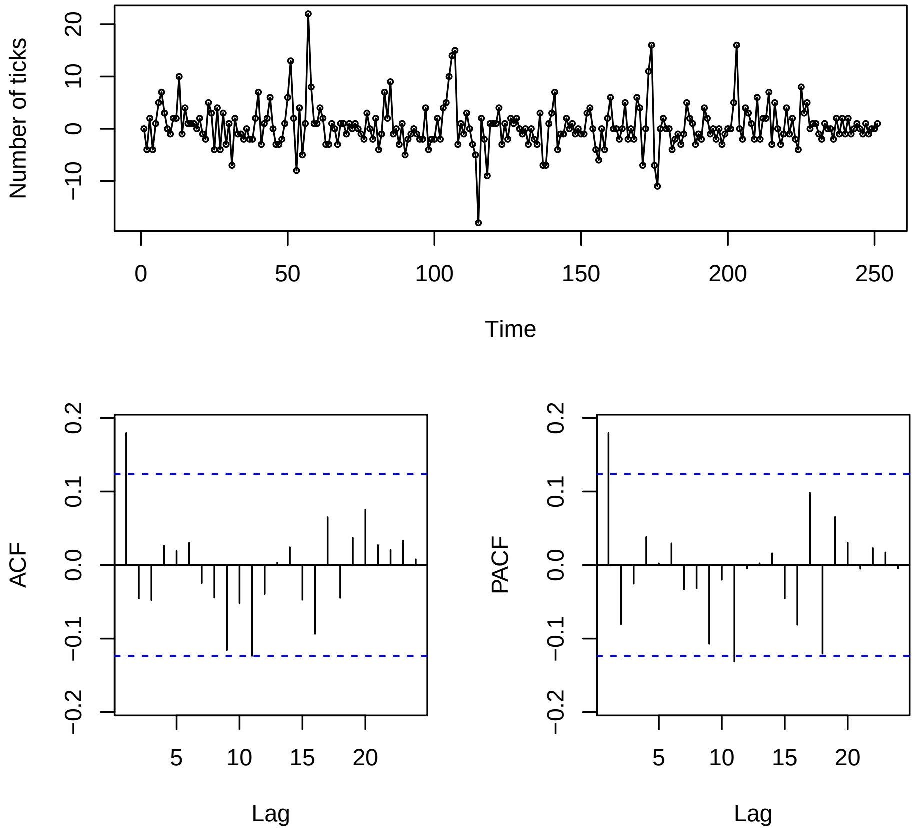


Figure 5.4: Time series plot, autocorrelation and partial autocorrelation functions for the number of ticks of the Saudi Telecom in 2012.

Table 5.4: Estimate of the parameters and the goodness-of-fit statistics RMS, MA and MdA for the NSINAR(1), SINARZ(1) and TINAR(1) processes.

Model	Estimates	RMS	MA	MdA
NSINAR(1)	$\widehat{\alpha} = 0.1795$	**4.181261**	2.931514	1.92750
	$\widehat{\mu} = 3.4263$			
	$\widehat{\lambda} = 2.9004$			
SINARZ(1)	$\widehat{\alpha} = 0.1795$			
	$\widehat{\mu}_1 = 2.6386$	4.181284	2.931514	1.92750
	$\widehat{\mu}_2 = 2.1127$			
TINAR(1)	$\widehat{\alpha} = 0.1795$			
	$\widehat{\lambda}_1 = 7.6272$	4.181284	2.931514	1.92750
	$\widehat{\lambda}_2 = 7.1957$			

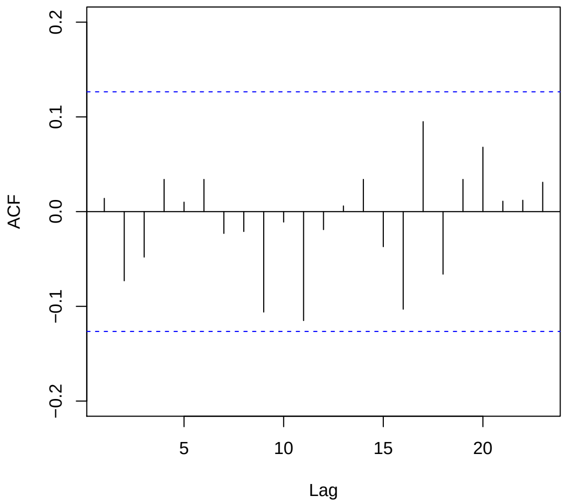


Figure 5.5: Sample autocorrelations of the residuals obtained from NSINAR(1) model.

5.7 Appendix

Proof of Proposition 5.

(i) is trivial.

(ii) follows from the fact that $1 \star Z$ is distributed as $1 * X - Y$, where X is geometrically distributed with mean μ, Y is Poisson(λ), X and Y independent. Then it suffices to verify that $1 * X$ is not distributed as X, which is easy to check. For example, $\Pr(1 * X = 0) = \mathrm{E}[\Pr(1 * X = 0|X)] = \mathrm{E}[2^{-X}] = 2/(2+\mu)$, while $\Pr(X = 0) = 1/(1+\mu)$.

(iii) and (iv) follow from the fact that $\alpha \star Z$ is distributed as $\alpha * X - \alpha \circ Y$, where X is geometrically distributed with mean μ, Y is Poisson(λ), X and Y independent.

(v) and (vi) will follow easily, since X and Y are independent.

(vii) can be deduced from the following general fact: let V be a discrete random variable such that $E|V| < \infty$ and let $\{A_k\}_{k\in\mathbb{Z}}$ be a sequence of random events with $\Pr(A_k) > 0$ for all k. Suppose that $\mathrm{E}[V|A_k]$ is a constant c not depending on k. Then, $\mathrm{E}[V| \cup A_k]$ is also equal to c.

Proof of Proposition 6.

We have

$$\begin{aligned}
\Pr(\alpha \star Z_{t-1} = j | Z_{t-1} = k) &= \frac{\Pr(\alpha \star Z_{t-1} = j, Z_{t-1} = k)}{\Pr(Z_{t-1} = k)} \\
&= \frac{1}{\Pr(Z_{t-1} = k)} \sum_{l=0}^{\infty} \Pr(\alpha \star Z_{t-1} = j, X_{t-1} = k + l, Y_{t-1} = l) \\
&= \frac{1}{\Pr(Z_{t-1} = k)} \sum_{l=0}^{\infty} \Pr(\alpha \star Z_{t-1} = j | X_{t-1} = k + l, Y_{t-1} = l) \\
&\times \Pr(X_{t-1} = k + l, Y_{t-1} = l) \\
&= \mathrm{e}^{-\frac{\lambda\mu}{1+\mu}} \sum_{l=0}^{\infty} \frac{[\lambda\mu/(1+\mu)]^l}{l!} \Pr(\alpha \star Z_{t-1} = j | X_{t-1} = k + l, Y_{t-1} = l).
\end{aligned} \tag{5.7}$$

We know that the random variables $\alpha * X_{t-1} | X_{t-1} = k + l$ and $\alpha \circ Y_{t-1} | Y_{t-1} = l$ have $\mathrm{NB}(k + l, \alpha/(1+\alpha))$ and $\mathrm{B}(l, \alpha)$ distributions. Thus,

$$\Pr(\alpha \star Z_{t-1} = j | X_{t-1} = k + l, Y_{t-1} = l) = \sum_{m=0}^{l} \binom{j + l + k - 1}{j} \binom{l}{m} \frac{\alpha^{j+m}(1-\alpha)^{l-m}}{(1+\alpha)^{l+k+j}}. \tag{5.8}$$

plugging (5.8) into (5.7), we obtain

$$\Pr(\alpha \star Z_{t-1} = j | Z_{t-1} = k) = \frac{\mathrm{e}^{-\frac{\lambda\mu}{1+\mu}} \alpha^j}{(1+\alpha)^{k+j}} \sum_{l=0}^{\infty} \sum_{m=0}^{l} \binom{j + l + k - 1}{j} \binom{l}{m} \frac{\alpha^m (1-\alpha)^{l-m}}{(1+\alpha)^l} \frac{[\lambda\mu/(1+\mu)]^l}{l!}.$$

Proof of Proposition 7.

The strict stationarity of the NSTINAR(1) process $\{Z_t\}_{t\in\mathbb{Z}}$ follows from the fact that it is a Markov process of first order and that the random variables $\{Z_t\}_{t\in\mathbb{Z}}$ are identically distributed random variables. Now, we will prove ergodicity. Let $\{V(t)\}$ be all counting series of $\alpha \star Z_{t-1}$ in (5.5). Let $\sigma(U_1, U_2, \dots)$ denote the σ-field generated by the sequence of random variables $U_1, U_2, \dots$; from (5.5), we obtain

$$\sigma(Z_t, Z_{t-1}, \dots) \subset \sigma(\xi_t, V(t), \xi_{t-1}, V(t-1), \dots)$$

and hence

$$\bigcap_{t=0}^{-\infty} \sigma(Z_t, Z_{t-1}, \dots) \subset \bigcap_{t=0}^{-\infty} \sigma(\xi_t, V(t), \xi_{t-1}, V(t-1), \dots). \tag{5.9}$$

Since $\{V(t), \xi_t\}$ are independent sequences, then, from Kolmogorov's Zero-One law, any event in the tail σ-field denoted by (5.9) has probability 0 or 1. From Lemma 2 of Shiryaev (1996) (pp. 408), $\{Z_t\}_{t\in\mathbb{Z}}$ is ergodic.

Proof of Proposition 8.

Strong consistency of these estimators follows from ergodicity of the process. For the NSTINAR(1) process, we have that $\sum_{i=0}^{\infty} \alpha^i < \infty$. Thus, the sufficient condition of Theorem 1 of Silva and Silva

(2006) is satisfied for the NSTINAR(1) process. So, we obtain that the estimator $(\overline{Z}, \widehat{\gamma}(0), \widehat{\gamma}(1))^\top$ has asymptotic normal distribution with mean $(\mu - \lambda, \mu(1+\mu) + \lambda, \alpha[\mu(1+\mu)+\lambda])^\top$ and variance $T^{-1}\mathbf{\Sigma}$, where the matrix $\mathbf{\Sigma}$ is defined as in Theorem 1 of Silva and Silva (2006).

Now, we use the delta method: suppose $\boldsymbol{X} = (x_1, x_2, x_3)^\top$, $\boldsymbol{\eta}$ and $\mathbf{\Sigma}$ are such that $\sqrt{T}(\boldsymbol{X} - \boldsymbol{\eta}) \xrightarrow{d} \mathrm{N}\left(0, \mathbf{\Sigma}\right)$. Let $f = (f_1, f_2, f_3)^\top$ be a mapping from $\mathbb{R}^3$ to $\mathbb{R}^3$ where each f_i is differentiable at $\boldsymbol{\eta}$. Let $\mathbf{D}$ be the Jacobian matrix of f with respect to $\boldsymbol{X}$ Then, $\sqrt{T}(f(\boldsymbol{X}) - f(\boldsymbol{\eta})) \xrightarrow{d} \mathrm{N}\left(0, \mathbf{D}\mathbf{\Sigma}\mathbf{D}^\top\right)$.

For $f_1(x_1, x_2, x_3) = x_3/x_2$, $f_2(x_1, x_2, x_3) = \sqrt{x_2 + x_1 + 1} - 1$, $f_3(x_1, x_2, x_3) = \sqrt{x_2 + x_1 + 1} - x_1 - 1$ and $\boldsymbol{\eta} = (\mu - \lambda, \mu(1+\mu) + \lambda, \alpha[\mu(1+\mu)+\lambda])^\top$, we obtain

$$\begin{aligned}
\frac{\partial f_1(x_1,x_2,x_3)}{\partial x_1} &= 0, \frac{\partial f_1(x_1,x_2,x_3)}{\partial x_2} = -\frac{x_3}{x_2^2}, \frac{\partial f_1(x_1,x_2,x_3)}{\partial x_3} = \frac{1}{x_2},\\
\frac{\partial f_2(x_1,x_2,x_3)}{\partial x_1} &= \frac{1}{\sqrt{x_1+x_2+1}}, \frac{\partial f_2(x_1,x_2,x_3)}{\partial x_2} = \frac{1}{\sqrt{x_1+x_2}}, \frac{\partial f_2(x_1,x_2,x_3)}{\partial x_3} = 0,\\
\frac{\partial f_3(x_1,x_2,x_3)}{\partial x_1} &= \frac{1}{\sqrt{x_1+x_2+1}} - 1, \frac{\partial f_3(x_1,x_2,x_3)}{\partial x_2} = \frac{1}{\sqrt{x_1+x_2+1}}, \frac{\partial f_3(x_1,x_2,x_3)}{\partial x_3} = 0.
\end{aligned}$$

Thus, $\sqrt{T}\left(f\left(\overline{Z}, \widehat{\gamma}(0), \widehat{\gamma}(1)\right) - (\alpha, \mu, \lambda)^\top\right) \xrightarrow{d} \mathrm{N}\left(0, \mathbf{D}\mathbf{\Sigma}\mathbf{D}^\top\right)$, with

$$\mathbf{D} = \begin{pmatrix} 0 & -\frac{\alpha}{\mu(1+\mu)+\lambda} & \frac{1}{\mu(1+\mu)+\lambda} \\ \frac{1}{1+\mu} & \frac{1}{1+\mu} & 0 \\ -\frac{\mu}{1+\mu} & \frac{1}{1+\mu} & 0 \end{pmatrix}.$$

References

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

Alzaid, A.A. and Omair, M.A. (2010). On the poisson difference distribution inference and applications. *Bull. Malays. Math. Sci. Soc.*, **33**, 17-45.

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

Barreto-Souza, W. and Bourguignon, M. (2014). A skew INAR(1) process on $\mathbb{Z}$. Advances in Statistical Analysis. DOI 10.1007/s10182-014-0236-2.

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

Freeland, R.K. and McCabe, B.P.M. (2004). Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, **25**, 701-722.

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

Hwang, Y., Kim, J. and Kweon, I. (2007). Sensor noise modeling using the Skellam distribution: Application to the color edge detection. Cvpr, *2007 IEEE Conference on Computer Vision and Patern Recognition*, Minneapolis, MN, 1-8.

Kachour, M. and Truquet, L. (2011). A p-order signed integer-valued autoregressive (SINAR(1)) model. *Journal of Time Series Analysis*, **32**, 223-236.

Kachour, M. and Yao, J. F. (2009). First-order rounded integer-valued autoregressive (RINAR(1)) process. *Journal of Time Series Analysis*, **30**, 417-448.

Karlis, D. and Ntzoufras, I. (2006). Bayesian analysis of the difference of count data. *Statist. Med.*, 25, 1885-1905.

Karlis, D. and Ntzoufras, I. (2009). Bayesian modeling of football outcomes: Using the Skellam's distribution for the goal difference. *IMA J. Manage. Math.*, **20**, 133–145.

Kim, H.Y., Park, Y. (2008). A non-stationary integer-valued autoregressive model. *Statistical Papers*, **49**, 485-502.

Nastić, A.S. and Ristić, M.M. (2012). Some geometric mixed INAR models. *Statistics and Probability Letters*, **82**, 805-811.

Nastić, A.S., Ristić, M.M. and Bakouch, H.S. (2012). A combined geometric INAR(p) model based on negative binomial thinning. *Mathematical and Computer Modelling*, **55**, 1665-1672.

Ong, S.H., Shimizu, K. and Ng, C.M. (2008). A class of discrete distributions arising from difference of two random variables. *Comput. Stat. Data Anal.*, **52**, 1490-1499.

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.

Silva, I and Silva, M.E. (2006). Asymptotic distribution of the Yule-Walker estimator for INAR(p) processes. *Statistics and Probability Letters*, **76**, 1655-1663.

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

Shiryaev, A.N. (1996). Probability, 2nd Edition, Springer.

Steutel, F.W. and van Harn, K. (1979). Discrete Analogues of Self-Decomposability and Stability. *Annals of Probability*, **7**, 893-899.

Zhang, H., Wang, D. and Zhu, F. (2010). Inference for INAR(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference*, **140**, 667-683.

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