Federal University of Pernambuco
Centre for Natural and Exact Sciences
Department of Statistics

Graduate Program in Statistics

FORECASTING TIME SERIES WITH INTEGER VALUES

Luz Milena Zea Fernández

Doctoral thesis

Recife
December 2013

Federal University of Pernambuco
Centre for Natural and Exact Sciences
Department of Statistics

Luz Milena Zea Fernández

FORECASTING TIME SERIES WITH INTEGER VALUES

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
December 2013

Catalogação na fonte
Bibliotecária Jane Souto Maior, CRB4-571

Zea Fernández, Luz Milena
Forecasting time series with integer values / Luz Milena Zea Fernández. - Recife: O Autor, 2013.
139 f., il., fig., tab.

Orientador: Klaus Leite Pinto Vasconcellos.
Tese (doutorado) - Universidade Federal de Pernambuco. CCEN, Estatística, 2013.

Inclui referências e apêndice.

1. Estatística Matemática. 2. Probabilidade. 3. Séries temporais. I. Vasconcellos, Klaus Leite Pinto (orientador). II. Título.

519.9 CDD (23. ed.) MEI2013 – 168

Universidade Federal de Pernambuco
Pós-Graduação em Estatística

22 de novembro de 2013

Nós recomendamos que a tese de doutorado de autoria de

Luz Milena Zea Fernández

intitulada

"*Forecasting Time Series with Integer Values*"

seja aceita como cumprimento parcial dos requerimentos para o grau de Doutor(a) em Estatística.

Coordenador da Pós-Graduação em Estatística

Banca Examinadora:

Klaus Leite Pinto Vasconcellos orientador

Francisco José de Azevedo Cysneiros

Audrey Helen Mariz de Aquino Cysneiros

Pedro Alberto Morettim (USP)

Valdério Anselmo Reisen (UFES)

Este documento será anexado à versão final da tese.

Dedico esta tese aos meu pais,
Dilma e Alvaro,
com muito amor e carinho.

Agradecimentos

São muitas as pessoas a quem gostaria de agradecer e o trabalho de listar a todos de uma maneira completa e justa parece-me infinito! Mesmo assim, seguem alguns agradecimentos.

Ao meu orientador, professor Klaus Vasconcellos, pela excelente orientação, por tudo o que me ensinou, pelo apoio e paciência que sempre me dedicou e que foram fundamentais para a conclusão deste trabalho.

Aos meus pais Alvaro e Dilma, minha irmã Ximena e meu sobrinho Alejandro, por me acompanhar, me apoiar, me acolher nos momentos mais difíceis e vibrar comigo a cada conquista. Mesmo estando longe eles sempre conseguiram transmitir-me todo seu amor e carinho, por isso minha gratidão para com eles é sem limites.

A todos os professores do programa de Pós-graduação em Estatística da UFPE por todos os ensinamentos, pelo respeito, atenção e seriedade que sempre dedicam a todos os alunos. Em particular, gostaria de agradecer aos professores e amigos Raydonal, Patrícia, Getúlio e Sérgio por todo apoio, companheirismo e amizade. Ao professor Cribari por seu excelente trabalho á frente da Pós-graduação e por encorajar meu retorno ao doutorado. Ao professor Gauss pela oportunidade de trabalho em sua linha de pesquisa.

A todos os colaboradores do programa de Pós-graduação em Estatística da UFPE. Em particular agradeço a Valéria pela forma prestativa e atenciosa com que conduz seu trabalho e pelo carinho, atenção e amizade que sempre me dedicou. Ao Clebson pelos muitos almoços divertidos que compartilhamos e pelas inúmeras vezes que me ajudou quando solicitei. Ao Lódino por sua alegria e tranquilidade contagiantes, que sempre estiveram presentes em nossas conversas. Ao Jymmy, por sua amizade, pelos nossos longos papos, por todos os cafés e por resolver eficientemente todos os problemas de informática da sala de doutorado, mesmo que sua resposta sempre comece com *é um problema geral...*

Aos colegas de doutorado e mestrado: Fernanda, Rodrigo, Marcelo, Laércio, Waldemar, Fernando, Dora, Sébastien, William, Tatiane, Tarciana, Abraão, Manoel, Artur, Heloísa, Gustavo, Daniel e Thiago, com os quais pude compartilhar, além de crescimento profissional, uma amizade sincera. Aos grandes amigos Frank, Manoel da Matemática Computacional e a minha amiga Edleide pelos bons momentos que tivemos.

As minhas vizinhas Giselda, Brigida e Juliana, pela atenção, pelos momentos divertidos que tivemos, pelas conversas e a Ju pela sua alegria e pelo carinho. A minha amiga Luz Marina por se preocupar comigo e me acompanhar mesmo de longe. Aos meus amigos Diego e Rebeca que me brindaram muita atenção, carinho e amizade. Ao amigo Nicolás pelo apoio incondicional de sempre.

Um agradecimento especial à minha querida amiga Laura, por chegar ao Recife no momento em que mais precisei, por todas nossas conversas, por todas nossas saídas, por todo o carinho que sempre me ofereceu, por toda a preocupação que sempre demonstrou e por sempre dar-me o conselho que eu precisava escutar. Laura você já faz parte de minha família.

Agradeço a todos que não mencionei aqui, mas que de alguma forma estiveram comigo nesta etapa de minha vida e foram igualmente importantes na conclusão do meu doutorado. A todos o meu muito obrigada!

Aos membros da banca examinadora pelas suas valiosas sugestões que com certeza contribuíram ao enriquecimento deste trabalho, especialmente ao professor Valdério Anselmo Reisen.

Agradeço à CAPES pelo apoio financeiro.

O que prevemos raramente ocorre;
o que menos esperamos geralmente acontece

–Benjamin Disraeli

Resumo

O estudo das séries temporais é um dos tópicos mais importantes da Estatística, tendo como propósito principal o desenvolvimento de métodos para modelagem de dados que exibem correlação ao longo do tempo. Tais modelos nos permitem fazer previsões. Dentro desta área, séries temporais de valores inteiros têm chamado a atenção dado que podem ser observadas em muitos contextos, por exemplo, o número de acidentes mensais em uma fábrica, ou, o número de peixes capturados em una determinada área do mar cada semana. Nas últimas três décadas tem aumentado o interesse em propor metodologias para estudar séries temporais de valores inteiros, especificamente métodos para obter previsões futuras, as quais devem ser números inteiros não negativos, devido à natureza discreta destas séries.

Neste trabalho concentramo-nos em propor e estudar novos procedimentos para fazer previsões nos chamados processos autorregressivos de primeira ordem de valores inteiros, Integer-valued first-order Autoregressive Process (INAR(1)) com distribuição marginal Poisson, e nos processos autorregressivos de primeira ordem condicionalmente heteroscedásticos de valores inteiros, Integer-valued first-order Autoregressive Conditional Heteroskedasticity Processes (INARCH(1)).

No processo INAR(1) fornecemos uma expressão analítica para o valor esperado da parte inteira da média condicional um passo à frente. Além disso, estudamos o comportamento de três preditores coerentes, ou seja, preditores que fornecem previsões de valor inteiro não negativo, considerando diferentes cenários e também estudamos seu poder preditivo considerando dados modelados incorretamente.

Na seção 4.2 apresentamos uma forma alternativa de definir o processo INARCH(1) baseado no operador thinning Poisson. Começamos o Capítulo 4 definindo o operador thinning Poisson e, a seguir, encontramos e provamos suas propriedades. Além disso, na seção 4.2 fornecemos uma demonstração da existência e unicidade da distribuição marginal do processo INARCH(1) e também apresentamos propriedades da distribuição condicional h passos à frente do processo INARCH(1). Adicionalmente na seções 4.4 and 4.5 propomos previsões um, dois e h passos à frente para o modelo INARCH(1). Dado que a distribuição condicional um passo à frente é uma distribuição Poisson, propomos sua moda e sua mediana estimadas como previsões um passo à frente. Embora a moda da distribuição Poisson tenha uma expressão analítica simples, não existe uma expressão para a mediana. Assim, usando estudos de simulação, não apresentados neste trabalho, nós propomos uma aproximação simples da mediana da distribuição Poisson, a qual tem bom desempenho em termos de erro quadrático médio e em termos de erro absoluto médio. Na seção 4.3 apresentamos propriedades tais como média e variância limites da mediana aproximada. Começamos demonstrando que a aproximação é fracamente condicionalmente consistente e a seguir conseguimos provar que dita aproximação é fortemente condicionalmente consistente. Também provamos que ela é fracamente consistente e a seguir apresentamos a demostração de que ela é fortemente consistente. Na seção 4.6 apresentamos uma distribuição que nos permite obter intervalos de previsão bilaterais e unilaterais. Nas seções 4.7, 4.8 e 4.9 apresentamos estudos de simulação de Monte Carlo que comparam os desempenhos dos preditores propostos.

Nas seções 3.4 e 4.10 ilustramos as metodologias de previsão estudadas e propostas com exemplos de dados reais que já têm sido estudados nos processos INAR(1) com marginal Poisson e INARCH(1) respectivamente.

Palavras-chave: Modelo INARCH(1); Operador thinning Poisson; Previsão; Séries temporais de valores inteiros; Simulação de Monte Carlo.

Abstract

The study of time series is one of the most important subjects in the statistical literature, the main purpose being to provide methods for modeling data sets that exhibit correlation over time and to allow to make predictions. Integer-valued time series have paid the attention because they occur in many contexts, for example, the numbers of accidents in a manufacturing plant each month, or the numbers of fishes caught in a particular area of sea each week, often as counts of events, objects or individuals in consecutive intervals or at consecutive points in time. In the last three decades, there has been an increasing interest in proposing methodologies to study integer-valued time series, including how to obtain non-negative and integer predictors.

We center our attention in studying and proposing new forecasting procedures for the Integer-valued first-order Autoregressive Process (INAR(1)) with Poisson marginal distribution, based on the binomial thinning operator and for the Integer-valued first-order Autoregressive Conditional Heteroskedasticity Process (INARCH(1)), which takes into account the overdispersion.

In Chapter 3 we provide an analytic expression for the expected value of the integer part of the one-step ahead conditional mean for the INAR(1) process. In addition, using Monte Carlo simulation, we present a study of the behavior of three coherent forecasts, i. e., predictors which to provide non-negative and integer valued forecasts, and, we also present a study of their predictive power under misspecified data.

In section 4.2, we present an alternative way to define the INARCH(1) process, based on the Poisson thinning operator. We begin Chapter 4 defining the Poisson thinning

operator and then we find and prove its properties. After that, we provide a proof of existence and uniqueness of the marginal stationary distribution of the INARCH(1) process and we present properties of its conditional distribution $h-$steps ahead. Additionally, in Section 4.4 and Section 4.5 we propose one, two and $h-$steps ahead forecasts for INARCH(1) processes. We propose the estimated mode and median of a Poisson distribution as forecasts one-step ahead. Although the mode of a Poisson distribution has an easy analytic expression, there is no expression for the median. Hence, by simulation study, which does not present in this work, we propose an easy approximation of the median of a Poisson distribution which works very well in terms of mean squared error and mean absolute error. In Section 4.3 we show properties such as mean and variance limits of the approximate median. We begin with the proof of weakly conditional consistency of the approximate median, and then we get to prove strongly conditional consistency. Further, we are able to demonstrate weakly consistency of the approximate median and then we get to prove its strongly consistency. In Section 4.6 we show a distribution which allows to obtain one-sided and two-sided predictions intervals. In sections 4.7, 4.8 and 4.9 we present Monte Carlo simulation studies that compare the behaviors of the proposed forecasts.

For the models considered here, in Section 3.4 and Section 4.10 we illustrate the proposed and studied approaches with different real data sets, which were studied in these processes.

Key-words: Forecasting; INARCH(1) processes; Integer-valued time series; Monte Carlo simulation; Poisson thinning operator.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Resumo

O estudo das séries temporais é um dos mais importantes tópicos da Estatística, tendo como propósito principal o desenvolvimento de métodos para modelagem de dados que exibem correlação ao longo do tempo. Tais modelos nos permitem fazer previsões. Nas últimas décadas, tem havido um crescente interesse no estudo de séries temporais em que os dados são números inteiros de tamanho não muito grande. Tais séries temporais de valores discretos podem ser observadas em muitos contextos. McKenzie [2003] mostra alguns exemplos: número de acidentes que ocorrem por mês em um local de trabalho, número de pacientes tratados por hora em uma unidade de emergência em um hospital, número de peixes capturados por semana em uma determinada área, número de linhas telefônicas ocupadas numa rede a cada meia hora, número de máquinas que estão em pleno funcionamento em um grande laboratório no início de cada dia.

Neste trabalho concentramo-nos nos chamados processos autorregressivos de primeira ordem de valores inteiros, Integer-valued first-order Autoregressive Process (INAR(1)), e nos processos autorregressivos de primeira ordem condicionalmente heteroscedásticos de valores inteiros, Integer-valued first-order Autoregressive Conditional Heteroskedasticity Processes (INARCH(1)). O processo INAR(1) é baseado no

operador thinning binomial, é considerado o análogo discreto do conhecido processo contínuo autorregressivo de primeira ordem, ou processo AR(1), e pode ter diferentes distribuições marginais, incluindo a distribuição Poisson, a distribuição Binomial Negativa e a distribuição Poisson generalizada (Weiß [2008]), sendo a distribuição Poisson a mais usada. Neste caso o processo é conhecido como o processo Poisson INAR(1). Na presença de sobredispersão, ou seja, a variância maior que a média, uma alternativa ao processo Poisson INAR(1) é o processo INARCH(1) que leva em conta a sobredispersão.

Um assunto muito importante nestes processos é como obter previsões coerentes, ou seja, previsões inteiras e não negativas. O propósito desta tese é estudar previsão nos processos Poisson INAR(1) e INARCH(1). No capítulo 3 comparamos por simulação de Monte Carlo os comportamentos das previsões um passo à frente propostas por Freeland [1998] para o processo INAR(1) e estudamos o poder preditivo do modelo sob modelagem incorreta dos dados. Nossa principal contribuição neste trabalho é o estudo de previsão que fizemos para o modelo INARCH(1). Começamos o capítulo 4 definindo o operador thinning Poisson. Na seção 4.2 apresentamos uma forma alternativa de definir o processo INARCH(1) baseado no operador thinning Poisson. Na seção 4.3 fornecemos uma prova da existência e unicidade da distribuição estacionária marginal do processo INARCH(1), apresentamos expressões analíticas para o $r-$ésimo momento ordinário, para a média e variância condicional $h-$passos à frente e para a função geradora de probabilidades condicional $h-$passos à frente do processo INARCH(1). Além disso, para o caso particular $h = 2$, encontramos uma expressão simples para a função de probabilidade condicional dois passos à frente. Dado que a distribuição condicional um passo à frente do processo INARCH(1) é uma distribuição Poisson, nós propomos sua média e sua moda como previsões um passo à frente na seção 4.4. Embora a moda de uma distribuição Poisson tenha uma expressão analítica simples, não existe uma expressão para sua mediana. Portanto, usando estudos de simulação não apresentados aqui, nós propomos uma aproximação simples da mediana da distribuição Poisson, que tem um bom desempenho em termos de erro quadrático médio e de erro absoluto médio. Na seção 4.3 apresentamos propriedades tais como média e variância limites da mediana aproximada. Começamos provando consistência condicional fraca da mediana aproximada estimada e então conseguimos provar sua consistência condicional forte. Também, provamos consistência fraca da mediana aproximada estimada e então conseguimos provar sua consistência forte. Embora não tenhamos uma expressão analítica viável para a função de probabilidade condicional $h-$passos à frente para $h \geqslant 3$, na seção 4.5 propomos uma forma recursiva de obter a previsão $h-$passos à frente para $h \geqslant 2$. Na seção 4.6 apresentamos uma distribuição que nos permite obter intervalos de previsão unilaterais e bilaterais. Nas seções 4.7, 4.8 e 4.9 apresentamos simulações de Monte Carlo que comparam os desempenhos das previsões propostas.

Nas seções 3.4 e 4.10 ilustramos os métodos propostos e estudados para os processos INAR(1) e INARCH(1), usando diferentes conjuntos de dados reais que já têm sido considerados nestes processos.

Este trabalho está divido da seguinte maneira: no capítulo 2 apresentamos os processos INAR(1) e INARCH(1) junto com suas principais propriedades. Nos capítulos 3 e 4 apresentamos as contribuições na previsão dos modelos INAR(1) e INARCH(1), respectivamente. Finalmente, no capítulo 5 mostramos as principais contribuições desta tese e os tópicos de pesquisa futuros.

Initial presentation

The study of time series is one of the most important subjects in the statistical literature, the main purpose being to provide methods for modeling data sets that exhibit correlation over time and to allow to make predictions. In the last three decades, there has been an increasing interest in studying and proposing methods to model and make forecasts of integer-valued time series, i.e., series taking values on the set of non-negative integers.

Integer-valued time series occur in many contexts, often as counts of events, objects or individuals in consecutive intervals or at consecutive points in time. McKenzie [2003] shows some examples: the numbers of accidents in a manufacturing plant each month, the numbers of patients treated by a hospital's accident and emergency unit each hour, the numbers of fishes caught in a particular area of sea each week, the numbers of busy lines in a telephone network noted every thirty minutes, and the numbers of lifts in a tall office building which are fully operational at the start of business each day. More examples can be found in Weiß [2008] and Weiß [2010].

Integer-valued time series with large values may still be analyzed by using continuous-valued time series which are normally distributed. This is reasonable because many common distributions for count data, such as binomial, Poisson and negative binomial, have an approximate normal distribution when the distribution mean is large. However, when the discrete time series have small values, the classical methodology is not appropriate. Notice that the simple procedure of multiplying an integer-valued random variable by a real constant not necessarily leads to an integer-valued random variable. Then, an alternative is to replace the multiplication by a random operation which allows to obtain an integer random variable. Such operation was introduced by Steutel and Harn [1979] and is called binomial thinning operation.

The Integer-valued first-order Autoregressive Process (INAR(1)) is based on the binomial thinning operator and is considered the discrete analogous of the known continuous first-order autoregressive process, or AR(1) process. The INAR(1) process can have different marginal distributions, including the Poisson, negative binomial and the generalized Poisson distributions (Weiß [2008]), the Poisson marginal distribution being the most commonly used. In the presence of overdispersion, i.e., variance greater than the mean, an alternative to the Poisson INAR(1) process is the Integer-valued first-order Autoregressive Conditional Heteroskedasticity Process (INARCH(1)), which is a special case of the *Autoregressive Conditional Poisson* (ACP) models introduced by Heinen [2003] and since they are closely related to classical GARCH(p, q) models, Ferland et al. [2006] suggested to refer to these models as INGARCH(p, q) models. Weiß [2010] referred to the INGARCH($p, 0$) models as INARCH(p) models. The INARCH(1) process takes into account the overdispersion.

A very important subject in these processes is how to obtain coherent forecasts, i. e., non-negative and integer forecasts. Some researchers studied forecasting for the INAR(1) process, the baseline work being the research of Freeland [1998]. However, to the best of our knowledge, forecasting in INARCH(1) processes has not been studied.

The aim of this thesis is to study forecasting in INAR(1) and INARCH(1) processes. In

Chapter 3 we compare by Monte Carlo simulation the behaviors of the one-step ahead forecasts proposed by Freeland [1998] for the INAR(1) process with Poisson marginal distribution and we study the predictive power of this process modeled under mis-specified data. However, our principal contributions are to the INARCH(1) process. We begin Chapter 4 defining the Poisson thinning operator and then we find and prove its properties. In section 4.2, we present an alternative way to define the INARCH(1) process, based on the Poisson thinning operator. In Section 4.3 we provide a proof of existence and uniqueness of the marginal stationary distribution of the INARCH(1) process, we obtain analytic expressions for the $r-$th marginal ordinary moment, for the $h-$steps conditional mean and variance as well as for the $h-$steps ahead conditional probability generating function. Besides, for the particular case $h = 2$, we find a simple expression for the two-steps ahead conditional probability function. Given that the INARCH(1) process has the advantage that the conditional distribution one-step ahead is a Poisson distribution, we propose its median and mode as forecasts one-step ahead in Section 4.4. Although the mode of a Poisson distribution has an easy analytic expression, there is no expression for the median. Hence, by simulation study, which does not present in this work, we propose an easy approximation of the median of a Poisson distribution which works very well in terms of mean squared error and mean absolute error. In Section 4.3 we show properties such as mean and variance limits of the approximate median. We begin with the proof of weakly conditional consistency of the approximate median, and then we get to prove strongly conditional consistency. Further, we are able to demonstrate weakly consistency of the approximate median and then we get to prove its strongly consistency. Although we obtain an analytical expression for the $h-$steps ahead conditional probability generating function in Section 4.3, it does not lead to a workable procedure to obtain the $h-$steps ahead conditional probability function for $h \geqslant 3$, so, in Section 4.5 we propose a recursive way to find the $h-$steps ahead forecast for $h \geqslant 2$. In Section 4.6 we show a distribution which allows to obtain one-sided and two-sided predictions intervals. In sections 4.7, 4.8 and 4.9 we present Monte Carlo simulation studies that compare the behaviors of the proposed forecasts.

For the models considered here, in Section 3.4 and Section 4.10 we illustrate the proposed and studied approaches with different real data sets, which were studied in these processes.

The outline of this thesis is as follows: in Chapter 2 the INAR(1) and INARCH(1) processes are defined and a review of the main properties so far obtained in the literature is carried out. In Chapter 3 and Chapter 4 we present our contribution to forecasting in INAR(1) and INARCH(1) processes, respectively. Finally, Chapter 5 refers to the main contributions of the thesis and topics that require further investigation.

Chapter 2

Integer valued autoregressive processes

Resumo

Neste capítulo apresentamos o modelo INAR(1) introduzido por McKenzie [1985], McKenzie [1988] e Al-Osh and Alzaid [1987], baseado no operador thinning binomial proposto por Steutel and Harn [1979], para modelar e gerar sequências de processos de valores inteiros dependentes. Além disso, mostramos as propriedades do operador thinning binomial junto com alguns métodos de estimação e previsão propostos por Freeland [1998] e Freeland and McCabe [2004] para o processo INAR(1). Apresentamos ainda o processo INARCH(1), que leva em conta a sobredispersão. Este processo é um caso especial do processo Poisson condicional autorregressivo *Autoregressive Conditional Poisson* (ACP) introduzido por Heinen [2003] e dado que ele está intimamente relacionados com o clássico modelo GARCH(p, q), Ferland et al. [2006] sugeriu chamar este processo como processo INGARCH(p, q). Weiß [2010] se refere aos processos INGARCH($p, 0$) como processos INARCH(p). Finalmente, apresentamos alguns métodos de estimação de parâmetros.

Initial presentation

In this chapter we present the INAR(1) model introduced by McKenzie [1985], McKenzie [1988] and Al-Osh and Alzaid [1987], which is based on the binomial thinning operator provided by Steutel and Harn [1979], for modeling and generating sequences of dependent counting processes. Also, we exhibit the binomial thinning operator properties together with some estimation and forecasting methods proposed in Freeland [1998] and Freeland and McCabe [2004] for the INAR(1) process. On the other hand, we present the Integer-valued first-order Autoregressive Conditional Heteroskedasticity process INARCH(1), which takes into account the overdispersion; this process is a special case of the *Autoregressive Conditional Poisson* (ACP) models introduced by Heinen [2003] and since they are closely related to classical GARCH(p, q) models, Ferland et al. [2006] suggested to refer to these models as INGARCH(p, q) models. Weiß [2010] referred to the INGARCH$(p, 0)$ models as INARCH(p) models. Finally, we exhibit some estimation methods.

2.1 The binomial thinning operator

Definition 1. *Let X be a non-negative integer-valued random variable and $\alpha \in [0,1]$. The thinning operator is defined by Steutel and Harn [1979] as follows:*

$$\alpha \circ X \stackrel{\mathrm{d}}{=} \sum_{i=1}^{X} N_i,$$

where N_i are independent and identically distributed (i.i.d) binary random variables, independent of X, with $P(N_i = 1) = \alpha = 1 - P(N_i = 0)$.

The notation $X \stackrel{\mathrm{d}}{=} Y$ means that X has the same distribution of Y. The sequence $N_1, N_2, \ldots$ is said to be the counting series of $\alpha \circ X$.

Note that given X, $\alpha \circ X$ has binomial distribution with parameters X and α, denoted as

$$\alpha \circ X \mid X \sim \text{Bin}(X, \alpha).$$

So, the operator '$\circ$' is known as *the binomial thinning operator*.

Different generalizations of the binomial thinning operator have been proposed by relaxing some of the assumptions of its definition. Latour [1998] defined the thinning operator such that the counting series $N_1, N_2, \ldots$ of $\alpha \circ X$ are i.i.d. and independent of X, but, in contrast to binomial thinning, they are now allowed to be any non-negative integer-valued random variables (with mean α and variance β). Brännäs et al. [2002] suggested to allow dependence between the random variables of the counting series $N_1, N_2, \ldots$ of $\alpha \circ X$. Kim and Park [2008] proposed the signed binomial thinning, allowing negative integers for the random variable X. For more details of the generalizations of the binomial thinning operator see Weiß [2008]. The properties of the binomial thinning operator are presented in the next lemma.

Lemma 1. *Let X, Y be non-negative integer-valued random variables. Let α, β be real constants in* [0, 1] *and suppose that the counting series of $\alpha \circ X$ is independent of the counting series of $\alpha \circ Y$ and independent of X and Y. Then*

i) $0 \circ X = 0$

ii) $1 \circ X = X$

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

iv) $\alpha \circ (X + Y) \stackrel{d}{=} \alpha \circ X + \alpha \circ Y$ *if X and Y are independent*

v) $\mathrm{E}[\alpha \circ X] = \alpha \, \mathrm{E}[X]$

vi) $\mathrm{Var}[\alpha \circ X] = \alpha^2 \, \mathrm{Var}[X] + \alpha(1-\alpha)\,\mathrm{E}[X]$

vii) $\mathrm{E}[\alpha \circ X \,|\, X] = \alpha X$

viii) $\mathrm{Var}[\alpha \circ X \,|\, X] = \alpha(1-\alpha)X$

ix) $\mathrm{E}\left[(\alpha \circ X)^2\right] = \alpha^2 \, \mathrm{E}[X^2] + \alpha(1-\alpha)\,\mathrm{E}[X]$

x) $\mathrm{E}\left[(\alpha \circ X)^3\right] = \alpha^3 \, \mathrm{E}[X^3] + 3\alpha^2(1-\alpha)\,\mathrm{E}[X^2] + \alpha(1-\alpha)(1-2\alpha)\,\mathrm{E}[X]$

xi) $\mathrm{E}[X(\alpha \circ Y)] = \alpha \, \mathrm{E}[XY]$

xii) $\mathrm{E}\left[X(\alpha \circ Y)^2\right] = \alpha^2 \mathrm{E}[XY^2] + \alpha(1-\alpha)\mathrm{E}[XY]$

xiii) $\mathrm{E}[(\alpha \circ X)(\beta \circ Y)] = \alpha\beta\mathrm{E}[XY]$

xiv) $\mathrm{E}\left[(\alpha \circ X)^2(\beta \circ Y)\right] = \alpha^2\beta\mathrm{E}[X^2Y] + \alpha(1-\alpha)\beta\mathrm{E}[XY]$

xv) $\mathrm{Cov}(\alpha \circ X, \beta \circ Y) = \alpha\beta\mathrm{Cov}(X, Y)$

The proofs of properties in Lemma 1 can be derived from Definition 1, using the well known formulae for conditional moments and probability generating functions; more details and other properties can be found in Silva and Oliveira [2004] and Silva [2005].

Weiß [2008] provided the following interpretation of the binomial thinning operator: consider a population of size X at a certain time t. If we observe the same population at a later point of time, say $t+1$, then, the population may be shrinked, because some of the individuals died between times t and $t+1$. If the individuals die independently of each other, and if the probability of dying in between t and $t+1$ is equal to $1-\alpha$ for all individuals, then the number of survivors is given by $\alpha \circ X$.

2.2 INAR(1) formulation model

Definition 2. *A discrete non-negative integer-valued process $(X_t)_{t\geqslant 1}$, $X_t \in \mathbb{N}_0$, is said to be an INAR(1) process if it satisfies the recursion*

$$X_t = \alpha \circ X_{t-1} + \epsilon_t \quad \text{for } t \geqslant 1, \tag{2.1}$$

where $\alpha \in [0,1]$, $\mathbb{N}_0 = \{0, 1, \dots\}$, '$\circ$' represents the thinning operator given in Definition 1, $(\epsilon_t)_{t\geqslant 1}$, $\epsilon_t \in \mathbb{N}_0$, is a sequence of non-negative integer-valued i.i.d. random variables with mean μ_ϵ and variance σ_ϵ^2 and it is assumed that the counting series of $\alpha \circ X_{t-1}$ is independent of ϵ_t.

From the first two properties of Lemma 1 note that for $\alpha = 1$, the process in the last definition is a random walk and for $\alpha = 0$ the process is the sequence $(\epsilon_t)_{t\geqslant 1}$. Further, Al-Osh and Alzaid [1987] and Du and Li [1991] proved that for $\alpha < 1$ a process $(X_t)_{t\geqslant 1}$

satisfying Equation 2.1 is a second-order stationary process, i.e., a process with constant mean and variance, and such that the covariance between X_t and X_{t+h} depends only on h, these processes are also known as weakly stationary processes or simply, stationary processes. In this thesis we assume $\alpha \in (0, 1)$ to guarantee stationarity of the INAR(1) process.

The INAR(1) process allows different marginal distributions, including the negative binomial and the generalized Poisson distribution (Weiß [2008]). However, the Poisson marginal distribution is the most commonly used. It is easy to show that if $(\epsilon_t)_{t\geqslant 1}$ have Poisson distribution with parameter λ, denoted as $\epsilon_t \sim \text{Po}(\lambda)$, and if $X_0 \sim \text{Po}(\lambda/(1-\alpha))$, then the process $(X_t)_{t\geqslant 1}$ satisfying the recursion equation 2.1, is a stationary Markov chain with marginal distribution $\text{Po}(\lambda/(1-\alpha))$ and transition probabilities $\pi_{ij} = \text{P}(X_t = j \mid X_{t-1} = i)$ given by

$$\pi_{ij} = \sum_{k=0}^{\min\{i,j\}} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} \exp(-\lambda) \frac{\lambda^{j-i}}{(j-i)!}.$$

Thus, if $\epsilon_t \sim \text{Po}(\lambda)$ then the process $(X_t)_{t\geqslant 1}$ satisfying the recursion equation 2.1, is known as Poisson INAR(1) process, the process considered in this thesis. The marginal mean and variance of the Poisson INAR(1) model are given by

$$\text{E}[X_t] = \text{Var}[X_t] = \frac{\lambda}{1-\alpha}. \tag{2.2}$$

From Equation 2.1 and using property *viii)* of Lemma 1 the conditional mean and variance of X_t given X_{t-1} for the Poisson INAR(1) process are respectively

$$\text{E}[X_t \mid X_{t-1}] = \alpha X_{t-1} + \lambda \quad \text{and} \quad \text{Var}[X_t \mid X_{t-1}] = \alpha(1-\alpha)X_{t-1} + \lambda.$$

Al-Osh and Alzaid [1987] showed that the autocovariance and autocorrelation functions of Poisson INAR(1) model at lag k can be expressed as

$$\begin{aligned} \gamma(k) &= \text{Cov}(X_{t-k}, X_t) = \alpha^k \gamma(0), \quad \text{for} \quad k = 0, 1, 2, \ldots, \\ \rho(k) &= \frac{\gamma(k)}{\gamma(0)} = \alpha^k, \quad \text{for} \quad k = 0, 1, 2, \ldots, \end{aligned}$$

respectively. Note that the autocorrelation function, $\rho(k)$, decays exponentially with lag k.

The INAR(1) model can be interpreted in different ways. In Freeland [1998] the model is interpreted as a birth and death process, i.e., each individual at time $t-1$ has probability α of continuing to be alive at time t, and, at each time t, the number of births follows a Poisson distribution with mean λ. Alternatively, he proposed to interpret the model as an infinite server queue for which the service time is geometric with parameter $1-\alpha$ and the arrival process is Poisson with mean λ. Moreover, the INAR(1) is also interpreted as a branching process with immigration, i.e., the outcome X_t is composed of the surviving elements of X_{t-1} during the period $(t-1, t]$, $\alpha \circ X_{t-1}$, and the number of immigrants during this period, ϵ_t. Each element of X_{t-1} survives with probability α and its survival has effect neither on the survival of the other elements nor on the number of immigrants, see Drost et al. [2008]. Then, we can interpret the INAR(1) model as follows

$$\underbrace{X_t}_{\text{Population at time } t} = \underbrace{\alpha \circ X_{t-1}}_{\text{Survivors of time } t-1} + \underbrace{\epsilon_t}_{\text{Immigrations}}.$$

2.3 Estimation methods for the Poisson INAR(1) model

Several estimation methods have been proposed for parameter estimation of the Poisson INAR(1) process. Jung et al. [2005] provided an extensive comparative study of the Yule-Walker, Generalized Method of Moments, Weighted Conditional Least Squares, Conditional Least Squares, Conditional Maximum Likelihood and Exact Maximum Likelihood estimators of the INAR(1) model. In this work, we considered three of them: Yule-Walker (YW) estimators, Conditional Least Squares (CLS) estimators and Conditional Maximum Likelihood (CML) estimators.

Let $X_1, X_2, \ldots, X_T$ be a time series generated according to the Poisson INAR(1) model, defined by recursion equation 2.1 with $\epsilon_t \sim \text{Po}(\lambda)$. The most simple approach to parameter estimation is the YW approach. Al-Osh and Alzaid [1987] proposed the first order sample autocorrelation as estimator for the parameter α, and an estimator of λ is based on the first moment of the Poisson marginal distribution (2.2), leading to

$$\widehat{\alpha}_{\text{YW}} = \frac{\sum_{t=2}^{T}(X_t - \overline{X}_T)(X_{t-1} - \overline{X}_T)}{\sum_{t=1}^{T}(X_t - \overline{X}_T)^2}, \quad \widehat{\lambda}_{\text{YW}} = (1 - \widehat{\alpha}_{\text{YW}})\overline{X}_T, \tag{2.3}$$

where $\overline{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$. Note that the $\widehat{\alpha}_{\mathrm{YW}}$ estimator can be negative and in this case the estimate of α is not coherent. Du and Li [1991] proved that the YW estimators are strongly consistent and Silva and Silva [2006] showed that $\widehat{\alpha}_{\mathrm{YW}}$ is asymptotically normally distributed.

Klimko and Nelson [1978] and Hall and Heyde [1980] considered CLS estimation for stochastic processes and as a particular case they obtained the parameter estimators of a stationary Markov process which coincide with the parameters of the Poisson INAR(1) process. The parameter estimates are chosen to minimize the sum

$$\sum_{t=2}^{T}\left[X_t - \mathrm{E}\left(X_t \mid \mathcal{F}_{t-1}\right)\right]^2,$$

where $\mathcal{F}_T$ denotes the $\sigma-$algebra generated by $\{X_t,\, 1 \leqslant t \leqslant T\}$. From Equation 2.1 note that given $X_1, \ldots, X_{t-1}$ we have that

$$\mathrm{E}\left(X_t \mid \mathcal{F}_{t-1}\right) = \alpha X_{t-1} + \lambda.$$

The resulting estimators of α and λ, the parameters of the Poisson INAR(1), can be derived explicitly and they have the following form

$$\widehat{\alpha}_{\mathrm{CLS}} = \frac{\sum_{t=2}^{T} X_t X_{t-1} - \frac{1}{T-1}\sum_{t=2}^{T} X_t \sum_{s=2}^{T} X_{s-1}}{\sum_{t=2}^{T} X_{t-1}^2 - \frac{1}{T-1}\left(\sum_{t=2}^{T} X_{t-1}\right)^2}, \quad \widehat{\lambda}_{\mathrm{CLS}} = \frac{1}{T-1}\left(\sum_{t=2}^{T} X_t - \widehat{\alpha}_{\mathrm{CLS}} \sum_{t=2}^{T} X_{t-1}\right). \tag{2.4}$$

Hall and Heyde [1980] and Al-Osh and Alzaid [1987] proved that $(\widehat{\alpha}_{\mathrm{CLS}}, \widehat{\lambda}_{\mathrm{CLS}})$ is strongly consistent estimator of (α, λ) as $T \longrightarrow \infty$ for all $(\alpha, \lambda) \in (0,1) \times (0, \infty)$. However the $\widehat{\alpha}_{\mathrm{CLS}}$ estimator can assume values outside the parametric space. Klimko and Nelson [1978] and Hall and Heyde [1980] showed asymptotic joint normality of $(\widehat{\alpha}_{\mathrm{CLS}}, \widehat{\lambda}_{\mathrm{CLS}})$. Further, using the theorem of Klimko and Nelson [1978], Freeland and McCabe [2005] provided a valid expression for the asymptotic covariance matrix of the CLS estimators, and it is given below

$$\sqrt{T}\begin{pmatrix} \widehat{\alpha}_{\mathrm{CLS}} - \alpha \\ \widehat{\lambda}_{\mathrm{CLS}} - \lambda \end{pmatrix} \xrightarrow{D} \mathcal{N}_2(\mathbf{0}, \Sigma_{\alpha,\lambda}),$$

with

$$\Sigma_{\alpha,\lambda} = \begin{pmatrix} 1-\alpha^2+\dfrac{\alpha}{\lambda}(1-\alpha)^2 & -\lambda(1+\alpha) \\ -\lambda(1+\alpha) & \lambda+\lambda^2\left(\dfrac{1+\alpha}{1-\alpha}\right) \end{pmatrix},$$

$\xrightarrow{D}$ denoting convergence in distribution and $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$ representing the bivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

Besides, Freeland and McCabe [2005] showed that the distribution of the CLS estimators is asymptotically equivalent to that of the estimators based on the Yule-Walker equations; thus neither is asymptotically more efficient than the other.

YW and CLS estimators are quite attractive for practice, because they have closed form and their asymptotic distribution is known. CML estimators are obtained by maximizing the logarithm of the conditional likelihood function and they usually need to be computed numerically using optimization methods. However Al-Osh and Alzaid [1987] showed that, as expected, CML estimators have less bias.

The unconditional likelihood function based on a sample $x_1, x_2, \ldots, x_T$ generated according to the Poisson INAR(1) process can be written as

$$\mathrm{L}(\alpha, \lambda; x_1, x_2, \ldots, x_T) = \frac{[\lambda/(1-\alpha)]^{x_1}}{x_1!} \exp[-\lambda/(1-\alpha)] \prod_{t=2}^{T} f_{X_t|X_{t-1}}(x_t \mid x_{t-1}; \alpha, \lambda),$$

where

$$f_{X_t|X_{t-1}}(x_t|x_{t-1}; \alpha, \lambda) = \exp(-\lambda) \sum_{k=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{k} \frac{\alpha^k(1-\alpha)^{x_{t-1}-k}\lambda^{x_t-k}}{(x_t-k)!}.$$

When x_1 is given, the conditional likelihood can be written as

$$\mathrm{L}(\alpha, \lambda; x_1, x_2, \ldots, x_T \mid x_1) = \prod_{t=2}^{T} f_{X_t|X_{t-1}}(x_t \mid x_{t-1}; \alpha, \lambda).$$

Thus, the conditional log-likelihood function is

$$\ell(\alpha, \lambda; x_1, x_2, \ldots, x_T \mid x_1) = \sum_{t=2}^{T} \log\big(f_{X_t|X_{t-1}}(x_t \mid x_{t-1}; \alpha, \lambda)\big).$$

The asymptotic distribution of the CML estimators was studied by Freeland and McCabe [2004] as well as Franke and Seligmann [1992], who showed that those estimators are asymptotically normally distributed. Additionally, Freeland and McCabe [2004] provided new expressions for the score function and the observed Fisher information, an explicit expression for the expected Fisher information of the complete time series is not yet available.

Freeland [1998] tested his estimation methods, namely, CLS, CML and GLS (generalized least squares) estimation methods, on some misspecified data. The author simulated 200 series of length 100 using binomial thinning with parameter $\alpha = 0.5$ and misspecifiying the arrival process by letting the distribution of ϵ_t be uniform over $\{0, 1, 2\}$. He compared the sampling distributions of $\widehat{\alpha}$ and $\widehat{\lambda}$ using his estimates and concluded that the CML estimates for α and λ are biased.

2.4 Forecasting for the Poisson INAR(1) model

Consider the problem to predict the future value x_{T+h} based on the observed series up to time T. A few researchers have investigated how to produce coherent forecasts for the Poisson INAR(1) model, i.e., how to produce integer-valued predictions. Silva et al. [2009] provided a Bayesian methodology to obtain integer-valued point predictions and Jung and Tremayne [2006] proposed a computer intensive method for generating integer predictions. On the other hand, Freeland [1998] considered two criteria for finding optimal forecasts, the minimum squared error and minimum absolute error of the forecasts. The squared error approach results in the conditional mean as the optimal forecast, while the absolute error approach yields the conditional median as the optimal forecast. The author considered the conditional mode as a third type of forecast, which is found by selecting the outcome with the largest probability. Thus, the $h-$steps ahead conditional distribution can be used to provide forecasts. Freeland [1998] showed the following theorem, which furnishes $h-$steps ahead conditional distribution for the Poisson INAR(1) model.

Theorem 2.1. *For the Poisson INAR(1) model defined by recursion equation 2.1 with $\epsilon_t \sim Po(\lambda)$, the distribution of X_{T+h} given X_T is the convolution of the binomial distribution with parameters α^h and X_T and a Poisson distribution with parameter $\lambda(1-\alpha^h)/(1-\alpha)$.*

That is, the h-step ahead conditional moment generating function is given by

$$\mathrm{M}_{X_{T+h}\,|\,X_T}(s) = \left[\alpha^h \exp(s) + (1-\alpha^h)\right]^{X_T} \exp\left[\lambda\left(\frac{1-\alpha^h}{1-\alpha}\right)(\exp(s)-1)\right], \quad s \in \mathbb{R},$$

where $\mathbb{R}$ denotes the set of the real numbers.

From the theorem above it is easy to see that the mean, variance and probability function of $X_{T+h}\,|\,X_T$ are given by

$$\begin{aligned} \mathrm{E}[X_{T+h}\,|\,X_T] &= \alpha^h X_T + \lambda\left(\frac{1-\alpha^h}{1-\alpha}\right), \quad h = 1, 2, \dots. \\ \mathrm{Var}[X_{T+h}\,|\,X_T] &= \alpha^h(1-\alpha^h)X_T + \lambda\left(\frac{1-\alpha^h}{1-\alpha}\right), \quad h = 1, 2, \dots. \end{aligned} \tag{2.5}$$

$$f_{X_{T+h}\,|\,X_T}(j\,|\,i\,;\,\alpha,\lambda) = \exp\left[-\lambda\left(\frac{1-\alpha^h}{1-\alpha}\right)\right] \sum_{k=0}^{\min\{i,j\}} \binom{i}{k} \left[\lambda\left(\frac{1-\alpha^h}{1-\alpha}\right)\right]^{j-k} \frac{(\alpha^h)^k(1-\alpha^h)^{i-k}}{(j-k)!}, \tag{2.6}$$

for $j = 0, 1, 2, \dots$.

Corollary 1. *Let $F_{X_{T+h}\,|\,X_T}$ denote the distribution function of $X_{T+h}\,|\,X_T$ and let F be the distribution function of a random variable W with Poisson distribution with mean $\frac{\lambda}{1-\alpha}$. Then*

$$X_{T+h}\,|\,X_T \xrightarrow{D} W \quad as \quad h \to \infty.$$

That is

$$\lim_{h\to\infty} F_{X_{T+h}\,|\,X_T}(x) = F(x),$$

for all x at which F is continuous.

From (2.5) as h goes to infinity the conditional mean and variance respectively go to the stationary unconditional mean and variance of the process. That is

$$\lim_{h\to\infty} \mathrm{E}\,[\,X_{T+h}\,|\,X_T\,] = \lim_{h\to\infty} \mathrm{Var}\,[\,X_{T+h}\,|\,X_T\,] = \frac{\lambda}{1-\alpha}.$$

It is easy to see that the forecast $X_T^{(\mathrm{Mc})}(h)$ which minimizes the mean squared error given the sample

$$\mathrm{E}\left[\left(X_{T+h} - X_T^{(\mathrm{Mc})}(h)\right)^2 \,\middle|\, X_T\right]$$

is the conditional mean $X_T^{(\mathrm{Mc})}(h) = \mathrm{E}\big[X_{T+h} \mid X_T\big]$. Its analytical expression is given in Equation (2.5).

Freeland [1998] proved that the forecast $X_T^{(\mathrm{Mn})}(h)$ which minimizes the expected absolute error given the sample

$$\mathrm{E}\Big[\Big|X_{T+h} - X_T^{(\mathrm{Mn})}(h)\Big| \,\Big|\, X_T\Big]$$

is the conditional median $X_T^{(\mathrm{Mn})}(h) = \min\Big\{m : \mathrm{F}_{X_{T+h} \mid X_T}(m \mid x_T; \alpha, \lambda) \geqslant \frac{1}{2}\Big\}$, where

$$\mathrm{F}_{X_{T+h} \mid X_T}(x \mid x_T; \alpha, \lambda) = \exp\left[-\lambda\left(\frac{1-\alpha^h}{1-\alpha}\right)\right] \sum_{l=0}^{x} \sum_{k=0}^{\min\{x_T\,,\,l\}} \binom{x_T}{k} \left[\lambda\left(\frac{1-\alpha^h}{1-\alpha}\right)\right]^{l-k} \times \frac{(\alpha^h)^k (1-\alpha^h)^{x_T-k}}{(l-k)!}$$

is the conditional distribution function of X_{T+h} given X_T.

Additionally, Freeland [1998] considered the conditional mode $X_T^{(\mathrm{Md})}(h)$ as a third type forecast. The conditional mode is the point at which the probability function of X_{T+h} given X_T is largest, i.e.,

$$X_T^{(\mathrm{Md})}(h) = \arg\max_x \Big\{f_{X_{T+h} \mid X_T}(x|x_T; \alpha, \lambda)\Big\},$$

where $f_{X_{T+h} \mid X_T}(x \mid x_T\,;\,\alpha, \lambda)$ is given in equation (2.6).

Although the conditional mean $X_T^{(\mathrm{Mc})}(h)$ is the only with analytical expression of the three forecasts considered by Freeland [1998], almost always it provides non-integer values. On the other hand, the conditional median $X_T^{(\mathrm{Mn})}(h)$ and the conditional mode $X_T^{(\mathrm{Md})}(h)$ are always integer values.

When a count-data time series exhibits overdispersion, i.e., the variance is greater than the mean, the marginal distribution is not able to be described by the Poisson INAR(1) model. Thus, it is necessary a model which takes into account the overdispersion. In the next section we present the Integer-valued first-order Autoregressive Conditional Heteroskedasticity Process INARCH(1), which takes into account the overdispersion; this process is a special case of the *Autoregressive Conditional Poisson* (ACP) models introduced by Heinen [2003] and since they are closely related to classical GARCH models, Ferland et al. [2006] suggested to refer to these models as INGARCH(p, q) models. Weiß [2010] referred to the INGARCH$(p, 0)$ models as INARCH(p) models.

2.5 INARCH(1) formulation model

The INGARCH(p, q) process with $p \geqslant 1$ and $q \geqslant 0$, p and q integers values, was defined by Ferland et al. [2006] as integer-valued analogue of the classical Generalized Autoregressive Conditional Heteroskedasticity (GARCH(p, q)) process.

Definition 3. *A discrete non-negative integer-valued process, $(X_t)_{t\geqslant 1}$, $X_t \in \mathbb{N}_0$, is called an INGARCH(p, q) process if it satisfies that*

$$\begin{cases} X_t \mid \mathcal{F}_{t-1} \sim \mathrm{Po}(\mu_t) \\ \mu_t = \lambda + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \delta_j \mu_{t-j} \end{cases}$$

where $\mathcal{F}_T$ is the sigma-field generated by $\{X_t,\, 1 \leqslant t \leqslant T\}$, $\lambda > 0$, $\alpha_i \geqslant 0$, $i = 1, \ldots, p$, $\delta_j \geqslant 0$, $j = 1, \ldots, q$.

For particular case $p = 1$ and $q = 0$ it is obtained the INGARCH(1, 0) process with $\mu_t = \alpha_1 X_{t-1} + \lambda$. Weiß [2010] referred to this process as the INARCH(1) process. Thus it is possible to define the INARCH(1) process as follows:

Definition 4. *A discrete non-negative integer-valued process, $(X_t)_{t\geqslant 1}$, $X_t \in \mathbb{N}_0$, is said to follow an INARCH(1) process if X_t conditioned on $\mathcal{F}_{t-1}$ is Poisson distributed with parameter $\mu_t = \alpha X_{t-1} + \lambda$, where $\lambda > 0$ and $\alpha \geqslant 0$.*

From definition above, it is easy to see that the conditional distribution of X_t conditioned on $X_{t-1}, \ldots, X_1$ is equidispersed, i. e.,

$$\mathrm{E}[X_t \mid \mathcal{F}_{t-1}] = \mathrm{Var}[X_t \mid \mathcal{F}_{t-1}] = \mu_t.$$

However the unconditional distribution shows overdispersion,

$$\begin{aligned} \mathrm{E}[X_t] &= \mathrm{E}[\mathrm{E}(X_t \mid \mathcal{F}_{t-1})] = \mathrm{E}[\mu_t], \\ \mathrm{Var}[X_t] &= \mathrm{E}[\mathrm{Var}(X_t \mid \mathcal{F}_{t-1})] + \mathrm{Var}[\mathrm{E}(X_t \mid \mathcal{F}_{t-1})] = \mathrm{E}[\mu_t] + \mathrm{Var}[\mu_t]. \end{aligned} \tag{2.7}$$

Given that $\mu_t = \alpha X_{t-1} + \lambda$ and using Equation 2.7, it is clear that the model is overdispersed whenever $\alpha > 0$ and the amount of overdispersion is an increasing function of α. In this thesis we assume $\alpha \in (0, 1)$ to guarantee stationarity and overdispersion of the INARCH(1) process.

Heinen [2003] and Ferland et al. [2006] demonstrated that the INARCH(1) process is a stationary process whenever $\alpha < 1$ with unconditional mean and variance given by

$$\mathrm{E}[X_t] = \frac{\lambda}{1-\alpha} \quad \text{and} \quad \mathrm{Var}[X_t] = \frac{\lambda}{(1-\alpha)(1-\alpha^2)}. \tag{2.8}$$

The autocorrelation function of the INARCH(1) model at lag k is expressed as

$$\rho(k) = Corr(X_{t-k}, X_t) = \alpha^k, \quad \text{for} \quad k = 0, 1, 2, \ldots .$$

Zhu and Wang [2011] showed that the INARCH(1) process has a unique stationary distribution and is uniformly ergodic. In this work, using an easy argument of Markov chains we prove that the stationary distribution of the INARCH(1) process exists and is unique. However, an explicit expression for the marginal distribution $\pi_j = \mathrm{P}\big(X_t = j\big)$ of an INARCH(1) process $(X_t)_{t \geqslant 1}$, $X_t \in \mathbb{N}_0$, is not known. Weiß [2010] analyzed two approaches to approximate the marginal process distribution:

✓ **First approach: Markov chain approximation**

Given that $(X_t)_{t \geqslant 1}$, $X_t \in \mathbb{N}_0$, has stationary distribution, it follows that

$$\pi_j = \lim_{n \to \infty} \pi_{ij}^n \quad \forall\, i, j \in \mathbb{N}, \quad \mathbb{N} = \{1, 2, 3, \ldots\}, \tag{2.9}$$

where the $n-$step transition probabilities π_{ij}^n

$$\pi_{ij}^n = \mathrm{P}\big(X_{t+n} = j \,|\, X_t = i\big) \tag{2.10}$$

follow recursively via

$$\pi_{ij}^n = \sum_{r=0}^{\infty} \pi_{ir} \pi_{rj}^{n-1}. \tag{2.11}$$

Equations 2.9, 2.10 and 2.11 allow to determine marginal probabilities numerically choosing $M, N \in \mathbb{N}$ sufficiently large, they approximate

$$\pi_j \approx \pi_{ij}^N, \quad \text{where} \quad \pi_{ij}^n \approx \sum_{r=0}^{M} \pi_{ir} \pi_{rj}^{n-1}$$

for arbitrary $i, j \in \mathbb{N}$.

✓ **Second approach: Poisson-Charlier expansion** The probability generating function (PGF) of X_t is defined as

$$\mathcal{G}_{X_t}(z) := \mathrm{E}\left[z^{X_t}\right]$$

and the factorial cumulant generating function (FCGF) of X_t is defined as

$$\widetilde{\kappa}_{X_t}(z) := \log\left[\mathcal{G}_{X_t}(1+z)\right] = \log\left\{\mathrm{E}\left[(1+z)^{X_t}\right]\right\};$$

the coefficients $\kappa_{(r)}$ of the series expansion $\widetilde{\kappa}_{X_t}(z) = \sum_{r=1}^{\infty} \frac{\kappa_{(r)}}{r!} z^r$ are referred to as *factorial cumulants*, with $\kappa_{(r)} = \frac{d^r \widetilde{\kappa}_{X_t}(z)}{dz}\Big|_{z=0}$.

Using the relation between the PGF and the FCGF,

$$\mathcal{G}_{X_t}(z) = \exp\left[\widetilde{\kappa}_{X_t}(z-1)\right] = \exp\left[\sum_{r=1}^{\infty} \frac{\kappa_{(r)}}{r!}(z-1)^r\right],$$

the author proposes to approximate the PGF of X_t as

$$\mathcal{G}_{X_t}(z) \approx \exp\left[\sum_{r=1}^{m} \frac{\kappa_{(r)}}{r!}(z-1)^r\right],$$

where the factorial cumulants $\kappa_{(r)}$ are obtained in terms of *cumulants* κ_r from the recursion equations

$$\kappa_{(1)} = \kappa_1 \quad \text{and} \quad \kappa_{(r)} = \alpha^r \kappa_r.$$

Weiß [2009] provided the following recursive way to obtain the cumulants of the INARCH(1) model. Let $\mathcal{M}_{X_t}(s)$ be the moment generating function of X_t, the cumulant generating function is defined as

$$\kappa_{X_t}(s) := \ln\big(\mathcal{M}_{X_t}(s)\big);$$

the coefficients κ_r of the series expansion $\kappa_{X_t}(s) = \sum_{r=1}^{\infty} \frac{\kappa_r}{r!} s^r$ are referred to as *cumulants*, with $\kappa_r = \frac{d^r \kappa_{X_t}(s)}{ds}\Big|_{s=0}$.

The first four cumulants for the INARCH(1) process are equal to

$$\kappa_1 = \mathrm{E}[X_t] = \zeta, \quad \kappa_2 = \mathrm{Var}[X_t], \quad \kappa_3 = \mathrm{E}[(X_t-\zeta)^3], \quad \kappa_4 = \mathrm{E}[(X_t-\zeta)^4] - 3\kappa_2^2.$$

The author demostrated that the cumulants of the INARCH(1) process can be determined recursively from

$$\kappa_1 = \frac{\lambda}{1-\alpha}, \quad \kappa_n = -(1-\alpha^n)^{-1} \sum_{j=1}^{n-1} s(n,j)\, \kappa_j \quad \text{for} \quad n \geqslant 2, \tag{2.12}$$

where $s(n,j)$ are the Stirling numbers of the first kind, i.e., the coefficients of x^j in the polynomial $(x)_n = x(x-1)\cdots(x-n+1)$,

$$(x)_n = \sum_{j=0}^{n} s(n,j) x^j,$$

the coefficients $s(n,j)$ are determined recursively by

$$s(n,0) = 0, \quad s(n,n) = 1 \quad \text{for} \quad n \geqslant 1,$$

$$s(n+1,j) = s(n,j-1) - n\, s(n,j) \quad \text{for} \quad j = 1,\ldots,n \quad \text{and} \quad n \geqslant 1.$$

From Equation 2.12 it is easy to see that

$$\kappa_1 = \frac{\lambda}{1-\alpha}, \quad \kappa_2 = \frac{\lambda}{(1-\alpha)(1-\alpha^2)},$$

$$\kappa_3 = \frac{1+2\alpha^2}{1-\alpha^3}\,\kappa_2 \quad \text{and} \quad \kappa_4 = \frac{1+6\alpha^2+5\alpha^3+6\alpha^5}{(1-\alpha^3)(1-\alpha^4)}\,\kappa_2.$$

Note that κ_1 and κ_2 coincide, respectively, with the unconditional mean and variance given in Equation 2.12.

2.6 Estimation methods for the INARCH(1) model

In this section we present the same three estimation methods considered for the Poisson INAR(1) model. The YW and CLS estimators for the INARCH(1) process have the same expressions than the YW and CLS estimators of the Poisson INAR(1) model; these expressions are given in formulas (2.3) and (2.4), respectively. Although the CLS estimators of α and λ have the same expressions for the Poisson INAR(1) and for the INARCH(1) models, and the parameters have a joint asymptotic normal distribution for both models, their asymptotic covariance matrices are different. Weiß [2010] showed that

$$\sqrt{T}\begin{pmatrix}\widehat{\alpha}_{\mathrm{CLS}}-\alpha\\ \widehat{\lambda}_{\mathrm{CLS}}-\lambda\end{pmatrix}\xrightarrow{D}\mathcal{N}_2(\mathbf{0},\Sigma_{\alpha,\lambda}),$$

where

$$\Sigma_{\alpha,\lambda}=\begin{pmatrix}\frac{1}{1-\alpha}\left[\lambda(1+\alpha)+\frac{1+2\alpha^4}{1+\alpha+\alpha^2}\right] & -\lambda(1+\alpha)-\frac{(1+2\alpha)\alpha^3}{1+\alpha+\alpha^2}\\ -\lambda(1+\alpha)-\frac{(1+2\alpha)\alpha^3}{1+\alpha+\alpha^2} & (1-\alpha^2)\left[1+\frac{\alpha(1+2\alpha^2)}{\lambda(1+\alpha+\alpha^2)}\right]\end{pmatrix}.$$

Let $x_1, x_2, \ldots, x_T$ be a sample generated according to the INARCH(1) process. The unconditional likelihood function can be expressed as

$$\begin{aligned}\mathrm{L}(\alpha,\lambda;x_1,\,x_2,\,\ldots,\,x_T)&=g_{X_t}(x_1)\prod_{t=2}^{T}g_{X_t|X_{t-1}}(x_t\mid x_{t-1};\alpha,\lambda)\\&=g_{X_t}(x_1)\prod_{t=2}^{T}\frac{(\alpha x_{t-1}+\lambda)^{x_t}\exp(-\alpha x_{t-1}-\lambda)}{x_t!};\end{aligned}$$

where $g_{X_t}(x_1)$ represents the probability function of the marginal distribution of X_t and $g_{X_t|X_{t-1}}(x_t\mid x_{t-1};\alpha,\lambda)$ denotes the probability function of the Poisson random variable with parameter $\alpha x_{t-1}+\lambda$.

On the other hand, the conditional log-likelihood function is given by

$$\begin{aligned}\ell(\alpha,\lambda;x_1,\,x_2,\,\ldots,\,x_T\mid x_1)&=\sum_{t=2}^{T}\log\big(g_{X_t|X_{t-1}}(x_t\mid x_{t-1};\alpha,\lambda)\big)\\&=\sum_{t=2}^{T}\big[x_t\log(\alpha x_{t-1}+\lambda)-\alpha x_{t-1}-\lambda-\log(x_t!)\big].\end{aligned}\tag{2.13}$$

From Equation 2.13 it is clear that the scores for α and λ are, respectively,

$$\begin{aligned}\frac{\partial\ell(\alpha,\lambda;x_1,\,x_2,\,\ldots,\,x_T\mid x_1)}{\partial\alpha}&=\sum_{t=2}^{T}\left(\frac{x_tx_{t-1}}{\alpha x_{t-1}+\lambda}-x_{t-1}\right),\\\frac{\partial\ell(\alpha,\lambda;x_1,\,x_2,\,\ldots,\,x_T\mid x_1)}{\partial\lambda}&=\sum_{t=2}^{T}\left(\frac{x_t}{\alpha x_{t-1}+\lambda}-1\right).\end{aligned}\tag{2.14}$$

The CML estimates of α and λ have to find by numerically maximizing $\ell(\alpha, \lambda; x_1, x_2, \ldots, x_T \,|\, x_1)$. The observed Fisher information $\mathcal{J}_{\alpha,\lambda}$ is obtained from Equation (2.14), $\mathcal{J}_{\alpha,\lambda}$ can be expressed as

$$\mathcal{J}_{\alpha,\lambda} = \begin{pmatrix} \sum_{t=2}^{T} \frac{x_t x_{t-1}^2}{(\alpha x_{t-1}+\lambda)^2} & \sum_{t=2}^{T} \frac{x_t x_{t-1}}{(\alpha x_{t-1}+\lambda)^2} \\ \sum_{t=2}^{T} \frac{x_t x_{t-1}}{(\alpha x_{t-1}+\lambda)^2} & \sum_{t=2}^{T} \frac{x_t}{(\alpha x_{t-1}+\lambda)^2} \end{pmatrix}.$$

An expression for the expected Fisher information $\mathcal{I}_{\alpha,\lambda} = \mathrm{E}[\mathcal{J}_{\alpha,\lambda}]$ is not available yet. Zhu and Wang [2011] found the asymptotic distribution for the CML estimators; the authors proved that

$$\sqrt{T} \begin{pmatrix} \widehat{\alpha}_{\mathrm{CML}} - \alpha \\ \widehat{\lambda}_{\mathrm{CML}} - \lambda \end{pmatrix} \xrightarrow{D} \mathcal{N}_2(\mathbf{0}, \mathcal{I}_{\alpha,\lambda}^{-1}).$$

Chapter 3

Forecasting for the Poisson INAR(1) model

Resumo

Neste capítulo estudamos as previsões propostas por Freeland [1998] no modelo INAR(1) e apresentadas na seção 2.4. O autor propõe a média, a mediana e a moda da distribuição condicional $h-$passos à frente como previsões $h-$passos à frente. Embora a média condicional seja a única das três previsões que tem expressão analítica, muitas vezes é um valor não inteiro, enquanto que a mediana e a moda condicionais são sempre valores inteiros não negativos. Começamos o capítulo apresentando uma expressão analítica para o erro quadrático médio das previsões $h-$passos à frente e a seguir concentramos nossa atenção nas previsões um passo à frente. Fornecemos a expressão analítica do valor esperado da parte inteira da média condicional um passo à frente e estudamos por simulação de Monte Carlo o comportamento de três previsões um passo à frente: parte inteira da média condicional, mediana condicional e moda condicional considerando parâmetros conhecidos e desconhecidos. No caso de parâmetros desconhecidos usamos os métodos de estimação Yule-Walker, mínimos quadrados condicionais e máxima verossimilhança condicional. Adicionalmente estudamos por simulação de Monte Carlo o poder preditivo do modelo Poisson INAR(1)

sob modelagem incorreta dos dados, modelamos incorretamente o processo de chegada como tendo distribuição uniforme discreta no conjunto $\{0, 1, 2, 3\}$. No final do capítulo ilustramos as metodologias estudadas com dois conjuntos de dados reais que já têm sido usados neste processo.

Initial presentation

In this chapter we study the forecasts for the Poisson INAR(1) process proposed by Freeland [1998] and presented in Section 2.4. The author proposed the mean, median and mode of the $h-$steps ahead conditional distribution as $h-$steps ahead forecasts. Although the conditional mean is the only of the three with analytical expression, it is most times a non-integer value while the conditional median and mode are always non-negatives integer values. We begin the chapter presenting an expression for the mean squared error of the forecasts $h-$steps ahead, after that, we focus on the one-step ahead forecasts. We provide an analytic expression for the expected value of the integer part of the one-step ahead conditional mean and we study by Monte Carlo simulation the behaviors of the three one-step ahead forecasts: integer part of the conditional mean, conditional median and conditional mode considering known and unknown parameters, for unknown parameters we use YW, CLS and CML estimation methods. Additionally we study by Monte Carlo simulation the predictive power of the Poisson INAR(1) model under misspecifed data, we misspecify the arrival process by letting its true distribution be uniform over $\{0, 1, 2, 3\}$. At the end of this chapter we illustrate the studied approaches with two different real data sets, which were studied in these processes.

3.1 Mean squared error of $h-$steps ahead forecasts

Given $X_1, X_2, \ldots, X_T$ a time series generated according to the Poisson INAR(1) process, Freeland [1998] proposed to predict a future value X_{T+h} by using $X_T^{(\mathrm{Mc})}(h)$, $X_T^{(\mathrm{Mn})}(h)$ and $X_T^{(\mathrm{Md})}(h)$, as presented in Section 2.4. From Equation 2.5 it is clear that the forecast

$X_T^{(\mathrm{Mc})}(h) = \mathrm{E}\left[X_{T+h} \,|\, X_T\right]$ almost always is a non-integer value. Thus, based on it we consider the integer-valued forecast

$$X_T^{(\mathrm{Ei})}(h) = \left\lfloor X_T^{(\mathrm{Mc})}(h) \right\rfloor = \left\lfloor \mathrm{E}\left[X_{T+h} |\, X_T\right] \right\rfloor = \left\lfloor \alpha^h X_T + \lambda \left(\frac{1-\alpha^h}{1-\alpha} \right) \right\rfloor,$$

where $\lfloor a \rfloor = \max\{m \in \mathbb{Z} : m \leqslant a\}$ is known as the floor of a or the integer part of a. Note that, for $h = 1$, $X_T^{(\mathrm{Ei})}(1) = \lfloor \alpha\, X_T + \lambda \rfloor$.

The integer-valued forecasts $X_T^{(\mathrm{Ei})}(h)$, $X_T^{(\mathrm{Mn})}(h)$ and $X_T^{(\mathrm{Md})}(h)$ can be compared by using the Mean Squared Error (MSE) or the Mean Absolute Error (MAE). Since they all depends on X_T, let $g(X_T)$ denote any of them, i. e., $g(X_T)$ is equal to $X_T^{(\mathrm{Ei})}(h)$, $X_T^{(\mathrm{Mn})}(h)$ or $X_T^{(\mathrm{Md})}(h)$. The MSE of $g(X_T)$ is given by

$$\mathrm{E}\left[\left(X_{T+h} - g(X_T) \right)^2 \right],$$

and the MAE of $g(X_T)$ is defined by

$$\mathrm{E}\left[\left| X_{T+h} - g(X_T) \right| \right]. \tag{3.1}$$

The next proposition provides an expression for the MSE of the forecasts $X_T^{(\mathrm{Ei})}(h)$, $X_T^{(\mathrm{Mn})}(h)$ and $X_T^{(\mathrm{Md})}(h)$.

Proposition 1. *If $g(X_T)$ denote $X_T^{(\mathrm{Ei})}(h)$, $X_T^{(\mathrm{Mn})}(h)$ or $X_T^{(\mathrm{Md})}(h)$, forecasts of X_{T+h}, then the MSE of $g(X_T)$ can be written as*

$$\mathrm{E}\left[\left(X_{T+h} - g(X_T) \right)^2 \right] = \frac{\lambda}{1-\alpha} + \left(\frac{\lambda}{1-\alpha} \right)^2 - 2\,\mathrm{E}\left\{ g(X_T) \left[\alpha^h \left(X_T - \frac{\lambda}{1-\alpha} \right) + \frac{\lambda}{1-\alpha} \right] \right\} + \mathrm{E}\left[g(X_T)^2 \right].$$

Proof. Let $g(X_T)$ be a forecast of X_{T+h} depending on X_T, $g(X_T)$ being equal to $X_T^{(\mathrm{Ei})}(h)$, $X_T^{(\mathrm{Mn})}(h)$ or $X_T^{(\mathrm{Md})}(h)$. Then the MSE of $g(X_T)$ can be written as

$$\begin{aligned}
\mathrm{E}\left[\left(X_{T+h} - g(X_T) \right)^2 \right] &= \mathrm{E}\left\{ \mathrm{E}\left[X_{T+h}^2 - 2X_{T+h}\, g(X_T) + g(X_T)^2 \,\middle|\, X_T \right] \right\} \\
&= \mathrm{E}\left\{ \mathrm{E}\left[X_{T+h}^2 | X_T\right] - 2g(X_T)\mathrm{E}\left[X_{T+h}|X_T\right] + g(X_T)^2 \right\} \\
&= \mathrm{E}\Big\{ \mathrm{Var}\left[X_{T+h}|X_T\right] + \mathrm{E}\left[X_{T+h}|X_T\right]^2 - 2g(X_T)\mathrm{E}\left[X_{T+h}|X_T\right]
\end{aligned}$$

$$
\begin{aligned}
&+ g(X_T)^2\Big\} \\
&= \mathrm{E}\left\{\alpha^h(1-\alpha^h)X_T + \lambda\left(\frac{1-\alpha^h}{1-\alpha}\right) + \left[\alpha^h\left(X_T - \frac{\lambda}{1-\alpha}\right) + \frac{\lambda}{1-\alpha}\right]^2\right. \\
&\quad \left. - 2g(X_T)\left[\alpha^h\left(X_T - \frac{\lambda}{1-\alpha}\right) + \frac{\lambda}{1-\alpha}\right] + g(X_T)^2\right\} \\
&= \alpha^h(1-\alpha^h)\mathrm{E}[X_T] + \lambda\left(\frac{1-\alpha^h}{1-\alpha}\right) + \alpha^{2h}\mathrm{E}\left[\left(X_T - \frac{\lambda}{1-\alpha}\right)^2\right] \\
&\quad + 2\frac{\lambda\alpha^h}{1-\alpha}\mathrm{E}\left[X_T - \frac{\lambda}{1-\alpha}\right] + \left(\frac{\lambda}{1-\alpha}\right)^2 \\
&\quad - 2\mathrm{E}\left\{g(X_T)\left[\alpha^h\left(X_T - \frac{\lambda}{1-\alpha}\right) + \frac{\lambda}{1-\alpha}\right]\right\} + \mathrm{E}\left[g(X_T)^2\right] \\
&= \frac{\lambda}{1-\alpha} + \left(\frac{\lambda}{1-\alpha}\right)^2 + -2\,\mathrm{E}\left\{g(X_T)\left[\alpha^h\left(X_T - \frac{\lambda}{1-\alpha}\right) + \frac{\lambda}{1-\alpha}\right]\right\} \\
&\quad + \mathrm{E}\left[g(X_T)^2\right].
\end{aligned}
$$

□

The forecast $X_T^{(\mathrm{Mc})}(1) = \alpha X_T + \lambda$ satisfies that

$$
\mathrm{E}\left[X_{T+1} - X_T^{(\mathrm{Mc})}(1)\right] = \frac{\lambda}{1-\alpha} - \mathrm{E}\left[\alpha X_T + \lambda\right] = \frac{\lambda}{1-\alpha} - \alpha\frac{\lambda}{1-\alpha} + \lambda = 0,
$$

but the same is not true for the forecast $X_T^{(\mathrm{Ei})}(1) = \lfloor \alpha X_T + \lambda \rfloor$, since

$$
\mathrm{E}\left[X_{T+1} - X_T^{(\mathrm{Ei})}(1)\right] = \frac{\lambda}{1-\alpha} - \mathrm{E}\left[\lfloor \alpha X_T + \lambda \rfloor\right].
$$

In the following proposition we present the expression to obtain $\mathrm{E}\left[\lfloor\, \alpha\ X_T\ +\ \lambda\ \rfloor\right]$.

Proposition 2. *Consider the forecast* $X_T^{(\mathrm{Ei})}(1) = \mathrm{E}\left[\lfloor \alpha X_T + \lambda \rfloor\right]$ *of* X_{T+1}. *Then its expected value can be expressed as*

$$
\mathrm{E}\left[\lfloor \alpha X_T + \lambda \rfloor\right] = \lfloor \lambda \rfloor \frac{\Gamma\left(\lceil a_{\lfloor \lambda \rfloor} + 1 \rceil, \mu\right)}{\left(\lceil a_{\lfloor \lambda \rfloor} + 1 \rceil - 1\right)!} + \sum_{m=\lfloor \lambda \rfloor + 1}^{\infty} m \left[\frac{\Gamma\left(\lceil a_{m+1} \rceil, \mu\right)}{\left(\lceil a_{m+1} \rceil - 1\right)!} - \frac{\Gamma\left(\lceil a_m \rceil, \mu\right)}{\left(\lceil a_m \rceil - 1\right)!}\right],
$$

where $a_m = (m-\lambda)/\alpha$, $\mu = \lambda/(1-\alpha)$, $\lceil a \rceil = \min\{m \in \mathbb{Z} : m \geqslant a\}$ *and* $\Gamma(a, x)$ *represents the incomplete gamma function, defined by* $\int_x^{\infty} s^{a-1}e^{-s}ds$.

Proof. The event $\Big[\lfloor \alpha X_T + \lambda \rfloor = m\Big]$ can be written as

$$\begin{aligned}
\Big[\lfloor \alpha X_T + \lambda \rfloor = m\Big] &= \Big[m \leqslant \alpha X_T + \lambda < m + 1\Big] \\
&= \left[\frac{m-\lambda}{\alpha} \leqslant X_T < \frac{m-\lambda+1}{\alpha}\right] \\
&= \Big[a_m \leqslant X_T < b_m\Big]
\end{aligned}$$

where $a_m = (m-\lambda)/\alpha$ and $b_m = (m-\lambda+1)/\alpha = a_m + 1/\alpha$. Depending on the values of α and λ, a_m and b_m can be integers or not, so we have the following possibilities

$$\Big[\lfloor \alpha X_T + \lambda \rfloor = m\Big] = \begin{cases} \Big[\lfloor a_m \rfloor \leqslant X_T \leqslant \lfloor b_m \rfloor - 1\Big], & \text{if} \quad a_m \in \mathbb{Z}, \quad b_m \notin \mathbb{Z} \\ \Big[\lfloor a_m \rfloor \leqslant X_T \leqslant \lfloor b_m \rfloor\Big], & \text{if} \quad a_m \in \mathbb{Z}, \quad b_m \in \mathbb{Z} \\ \Big[\lfloor a_m \rfloor + 1 \leqslant X_T \leqslant \lfloor b_m \rfloor - 1\Big], & \text{if} \quad a_m \notin \mathbb{Z}, \quad b_m \in \mathbb{Z} \\ \Big[\lfloor a_m \rfloor + 1 \leqslant X_T \leqslant \lfloor b_m \rfloor\Big], & \text{if} \quad a_m \notin \mathbb{Z}, \quad b_m \notin \mathbb{Z}. \end{cases}$$

Then,

$$\begin{aligned}
\Big[\lfloor \alpha X_T + \lambda \rfloor = m\Big] &= \Big[\lfloor a_m \rfloor - I_{\mathbb{Z}}(a_m) + 1 \leqslant X_T \leqslant \lfloor b_m \rfloor - I_{\mathbb{Z}}(b_m)\Big] \\
&= \Big[k(a_m) + 1 \leqslant X_T \leqslant k(b_m)\Big],
\end{aligned}$$

where $k(x) = \lfloor x \rfloor - I_{\mathbb{Z}}(x)$, i.e., $k(x)$ is the greatest integer less than x and I_A denotes the indicator function of A.

Using that $X_T \sim \mathrm{Po}(\mu)$, $\mu = \lambda/(1-\alpha)$, and the last equation, we obtain the following expression for the probability distribution of $\lfloor \alpha X_T + \lambda \rfloor$.

$$\begin{aligned}
\mathrm{P}\Big(\lfloor \alpha X_T + \lambda \rfloor = m\Big) &= \mathrm{P}\Big(k(a_m) + 1 \leqslant X_T \leqslant k(b_m)\Big) \\
&= \begin{cases} 0 & \text{if} \quad k(b_m) < 0, \\ \displaystyle\sum_{j=\max\{k(a_m)+1,0\}}^{k(b_m)} \left[\frac{\mu^j \exp(-\mu)}{j!}\right] & \text{if} \quad k(b_m) \geqslant 0. \end{cases}
\end{aligned}$$

Furthermore,

✓ If $m \leqslant \lfloor\lambda\rfloor - 2$, then $k(b_m) < 0$

$$m \leqslant \lfloor\lambda\rfloor - 2 < \lambda - 1 \quad \Rightarrow \quad m < \lambda - 1 \quad \Rightarrow \quad \frac{m-\lambda+1}{\alpha} < 0$$

$$\Rightarrow \quad \left\lfloor \frac{m-\lambda+1}{\alpha} \right\rfloor - I_{\mathbb{Z}}\left(\frac{m-\lambda+1}{\alpha}\right) \leqslant -1 \quad \Rightarrow \quad k(b_m) \leqslant -1 < 0$$

✓ If $m = \lfloor\lambda\rfloor - 1$, then $k(b_m) < 0$

$$m = \lfloor\lambda\rfloor - 1 \quad \Rightarrow \quad b_m = \frac{\lfloor\lambda\rfloor - \lambda}{\alpha} \leqslant 0 \quad \Rightarrow \quad k(b_m) = \lfloor b_m \rfloor - I_{\mathbb{Z}}(b_m) \leqslant 0$$

If $k(b_m) = 0$, then $\lfloor b_m \rfloor = I_{\mathbb{Z}}(b_m) = 0$, therefore $b_m \notin \mathbb{Z}, 0 \leqslant b_m < 1$ and $b_m \leqslant 0$; since this is a contradiction, thence $k(b_m) < 0$

✓ If $m = \lfloor\lambda\rfloor$, then $k(b_m) \geqslant 0$

$$m = \lfloor\lambda\rfloor \quad \Rightarrow \quad b_m = \frac{\lfloor\lambda\rfloor - \lambda + 1}{\alpha} > 0 \quad \Rightarrow \quad \lfloor b_m \rfloor \geqslant 0,$$

so, the only case in which $k(b_m) = \lfloor b_m \rfloor - I_{\mathbb{Z}}(b_m)$ could be negative is when $\lfloor b_m \rfloor = 0$ and $I_{\mathbb{Z}}(b_m) = 1$ and this happens if and only if $b_m = 0$; this implies $\lambda - \lfloor\lambda\rfloor = 1$ and this is impossible, thence $k(b_m) \geqslant 0$.

✓ If $m \geqslant \lfloor\lambda\rfloor + 1$, then $k(b_m) \geqslant 0$

$$\lfloor\lambda\rfloor + 1 > (\lambda - 1) + \alpha \quad \Rightarrow \quad m \geqslant (\lambda - 1) + \alpha \quad \Rightarrow \quad \frac{m-\lambda+1}{\alpha} \geqslant 1$$

$$\Rightarrow \quad \lfloor b_m \rfloor \geqslant 1 \quad \Rightarrow \quad k(b_m) = \lfloor b_m \rfloor - I_{\mathbb{Z}}(b_m) \geqslant 0$$

Therefore,

$$\mathrm{P}\Big(\lfloor \alpha X_T + \lambda \rfloor = m\Big) = \begin{cases} 0 & \text{if} \quad m \leqslant \lfloor\lambda\rfloor - 1, \\ \displaystyle\sum_{j=\max\{k(a_m)+1,0\}}^{k(b_m)} \left[\frac{\mu^j \exp(-\mu)}{j!}\right] & \text{if} \quad m \geqslant \lfloor\lambda\rfloor. \end{cases}$$

It is clear that

✓ $a_{m+1} = (m + 1 - \lambda)/\alpha = b_m$

✓ For $m \geqslant \lfloor\lambda\rfloor + 1$ we have $k(a_m) + 1 = \lceil a_m \rceil$

$$\checkmark\ k\big(a_{\lfloor\lambda\rfloor}\big) = k\left(\frac{\lfloor\lambda\rfloor-\lambda}{\alpha}\right) = \left\lfloor\frac{\lfloor\lambda\rfloor-\lambda}{\alpha}\right\rfloor - I_{\mathbb{Z}}\left(\frac{\lfloor\lambda\rfloor-\lambda}{\alpha}\right) \leqslant -1$$

Hence we can express the probability function of $\lfloor\alpha X_T+\lambda\rfloor$ as

$$\mathrm{P}\Big(\lfloor\alpha X_T+\lambda\rfloor = m\Big) = \begin{cases} 0 & \text{if} \quad m \leqslant \lfloor\lambda\rfloor-1, \\ \displaystyle\sum_{j=0}^{\lceil a_{\lfloor\lambda\rfloor+1}\rceil-1}\left[\frac{\mu^j\exp(-\mu)}{j!}\right] & \text{if} \quad m = \lfloor\lambda\rfloor, \\ \displaystyle\sum_{j=\lceil a_m\rceil}^{\lceil a_{m+1}\rceil-1}\left[\frac{\mu^j\exp(-\mu)}{j!}\right] & \text{if} \quad m \geqslant \lfloor\lambda\rfloor+1. \end{cases}$$

Then, the expected valued of $\lfloor\alpha X_T+\lambda\rfloor$ can be expressed as

$$\begin{aligned} \mathrm{E}\big[\lfloor\alpha X_T+\lambda\rfloor\big] &= \sum_{j=0}^{\lceil a_{\lfloor\lambda\rfloor+1}\rceil-1}\frac{\lfloor\lambda\rfloor\mu^j\exp(-\mu)}{j!} + \sum_{m=\lfloor\lambda\rfloor+1}^{\infty}\ \sum_{j=\lceil a_m\rceil}^{\lceil a_{m+1}\rceil-1}\frac{m\mu^j\exp(-\mu)}{j!} \\ &= \lfloor\lambda\rfloor\frac{\Gamma\big(\lceil a_{\lfloor\lambda\rfloor}+1\rceil,\mu\big)}{\big(\lceil a_{\lfloor\lambda\rfloor}+1\rceil-1\big)!} + \sum_{m=\lfloor\lambda\rfloor+1}^{\infty} m\left[\frac{\Gamma\big(\lceil a_{m+1}\rceil,\mu\big)}{\big(\lceil a_{m+1}\rceil-1\big)!} - \frac{\Gamma\big(\lceil a_m\rceil,\mu\big)}{\big(\lceil a_m\rceil-1\big)!}\right], \end{aligned}$$

where the last equation follows from the relation between the Poisson distribution function and the incomplete gamma function

$$\mathrm{F}(b) = \sum_{j=0}^{b}\frac{\mu^j\exp(-\mu)}{j!} = \frac{\Gamma(b+1,\mu)}{b!}.$$

□

3.2 Monte Carlo results for forecasting

In order to compare the performances of $X_T^{(\mathrm{Ei})}(1)$, $X_T^{(\mathrm{Mn})}(1)$ and $X_T^{(\mathrm{Md})}(1)$ as forecasts of X_{T+1}, we present some results of a Monte Carlo simulation that compares the square root of MSE, referred to as the Root Mean Square Error (RMSE), and the MAE for these forecasts, considering known and unknown parameters. We used 10000 Monte Carlo replications and considered different values of α and λ, namely, α = 0.1, 0.5, 0.9 and λ = 0.5, 1, 3, 5. The sample sizes considered are T = 25, 50 and 100.

In the first scenario we consider known parameters. We generate $R = 10000$ independent values of X_1 according to Po(μ) and then, given X_1, we generate X_2 using equation (2.1), i. .e.,

$$\begin{array}{ccc}
x_1^{(1)} = \mathrm{Po}(\mu) & \dashrightarrow & x_2^{(1)} = \mathrm{Bin}(x_1^{(1)}, \alpha) + \mathrm{Po}(\lambda) \\
\vdots & \vdots & \vdots \\
x_1^{(R)} = \mathrm{Po}(\mu) & \dashrightarrow & x_2^{(R)} = \mathrm{Bin}(x_1^{(R)}, \alpha) + \mathrm{Po}(\lambda)
\end{array}$$

where $\mu = \lambda/(1-\alpha)$.

Using Equation (3.1) and Proposition 1 with $h = 1$ we obtained the simulated MAE and the simulated RMSE of conditional median, conditional mode and integer part of conditional mean, as forecasts of X_2. The results are presented in table 3.1. Note that, for $\alpha = 0.1$ and $\alpha = 0.5$ the three forecasts were competitive in terms of RSME and MAE, and for $\alpha = 0.9$ the conditional median and conditional mode were competitive and they were a little better than the integer part of conditional mean in terms of RMSE and MAE.

In summary, when the parameters are known, the three forecasts: conditional median, conditional mode and integer part of the conditional mean were competitive, the integer part of conditional mean being a little worse for $\alpha = 0.9$, in terms of RMSE and MAE.

In the second scenario we consider unknown parameters. We simulated Monte Carlo samples X_1, X_2, ..., X_T, X_{T+1} and we estimated the parameters α and λ for each sample using the three estimation methods introduced in Section 2.3. For each sample, we verified if the parameter estimates are in parametric space; if it does not happen, then, we discarded the sample and substitute it by another Monte Carlo sample; for each valid sample we found the forecasts $X_T^{(\mathrm{Ei})}(1)$, $X_T^{(\mathrm{Md})}(1)$ and $X_T^{(\mathrm{Md})}(1)$ using the estimates of α and λ provided by each estimation method. After that, we calculated the simulated MAE and the simulated RMSE of the three forecasts.
Repeat until $r = 10000$

1. $x_1^{(r)} \sim \mathrm{Po}(\mu)$

	Error	Forecast	α		
			0.1	0.5	0.9
$\lambda = 0.5$	RMSE	$X_1^{(Mn)}(1)$	0.90	0.94	1.04
		$X_1^{(Md)}(1)$	0.93	0.94	1.04
		$X_1^{(Ei)}(1)$	0.93	0.94	**1.14**
	MAE	$X_1^{(Mn)}(1)$	0.55	0.63	0.68
		$X_1^{(Md)}(1)$	0.56	0.63	0.68
		$X_1^{(Ei)}(1)$	0.56	0.63	**0.77**
$\lambda = 1$	RMSE	$X_1^{(Mn)}(1)$	1.06	1.28	1.42
		$X_1^{(Md)}(1)$	1.06	1.28	1.42
		$X_1^{(Ei)}(1)$	1.06	1.28	**1.48**
	MAE	$X_1^{(Mn)}(1)$	0.77	0.94	1.04
		$X_1^{(Md)}(1)$	0.77	0.94	1.04
		$X_1^{(Ei)}(1)$	0.77	0.94	**1.10**
$\lambda = 3$	RMSE	$X_1^{(Mn)}(1)$	1.85	2.16	2.70
		$X_1^{(Md)}(1)$	1.86	2.16	2.70
		$X_1^{(Ei)}(1)$	1.86	2.16	**2.74**
	MAE	$X_1^{(Mn)}(1)$	1.42	1.66	1.87
		$X_1^{(Md)}(1)$	1.42	1.66	1.88
		$X_1^{(Ei)}(1)$	1.42	1.66	**1.90**
$\lambda = 5$	RMSE	$X_1^{(Mn)}(1)$	2.38	2.78	3.77
		$X_1^{(Md)}(1)$	2.41	2.78	3.77
		$X_1^{(Ei)}(1)$	2.41	2.78	**3.79**
	MAE	$X_1^{(Mn)}(1)$	1.86	2.17	2.44
		$X_1^{(Md)}(1)$	1.87	2.17	2.44
		$X_1^{(Ei)}(1)$	1.87	2.17	**2.46**

Table 3.1: RMSE and MAE of $X_1^{(Ei)}(1)$, $X_1^{(Mn)}(1)$ and $X_1^{(Md)}(1)$ for different values of known parameters α and λ.

2. $x_t^{(r)} = \text{Bin}(x_{t-1}^{(r)}, \alpha) + \text{Po}(\lambda), t = 2, \ldots, T+1$

   $\dashrightarrow \quad x_1^{(r)}, x_2^{(r)}, \ldots, x_T^{(r)}, x_{T+1}^{(r)}$

3. $x_1^{(r)}, x_2^{(r)}, \ldots, x_T^{(r)} \quad \dashrightarrow \quad (\widehat{\alpha}_{\text{YW}}, \widehat{\lambda}_{\text{YW}}), (\widehat{\alpha}_{\text{CLS}}, \widehat{\lambda}_{\text{CLS}}), (\widehat{\alpha}_{\text{CML}}, \widehat{\lambda}_{\text{CML}})$

4. If $(\widehat{\alpha}_{\text{YW}}, \widehat{\lambda}_{\text{YW}}), (\widehat{\alpha}_{\text{CLS}}, \widehat{\lambda}_{\text{CLS}}) \in (0,1) \times (0, \infty)$

   $\dashrightarrow \quad X_T^{(\text{Ei})}(1)^{(r)}$, $X_T^{(\text{Md})}(1)^{(r)}$ and $X_T^{(\text{Md})}(1)^{(r)}$; $\quad r = r+1$ and return to step 1

   Else

   $\dashrightarrow$ return to step 1 without calculating forecasts and without updating r.

Tables 3.2 and 3.3 show the RMSE and the MAE of these forecasts. For $\lambda = 0.5$, 1, $\alpha = 0.1$, 0.5 and for all sample sizes considered the three forecasts and the estimation methods considered were competitive, the conditional median being sightly better than the others, in terms of RMSE and MAE. For $\lambda = 0.5$, 1 $\alpha = 0.9$ and $T = 25$, YW estimators were worse than CLS and CML estimators, CML being slightly better than CLS, in terms of RMSE and MAE, while for $T = 50$ and $T = 100$ the conditional mean and mode were competitive and a little better than integer part of conditional mean, in terms of RMSE and MAE. On the other hand, for $\lambda = 3$, 5, $\alpha = 0.1$, 05, $T = 25$ and $T = 50$ the three forecasts and the estimation methods considered were competitive, while for $\lambda = 3$, 5, $\alpha = 0.9$ and $T = 25$, YW estimators were worse than CLS and CML estimators, CLS being slightly better than CML, in terms of RMSE and MAE. For $\lambda = 3$, 5, $\alpha = 0.9$ and $T = 100$ the estimation methods and the forecasts considered were competitive, in terms of RMSE and MAE.

In summary, for large values of α and small sample sizes YW estimators were worse than CLS and CML estimators, CLS being slightly better than CML, in terms of RMSE and MAE, for small and moderate sample sizes the conditional mean was slightly better than the others. For large sample sizes the three forecasts: conditional median, conditional mode and integer part of the conditional mean, and, the three estimation methods considered were competitive, in terms of RMSE and MAE. Therefore, we suggest to use the conditional median as forecast and CLS estimators, because they have explicit expression while CML are calculated using numerical maximization and CML estimators are only a little better than CLS estimators, in terms, of RMSE and MAE.

	Error	Forecast	Estimator	T = 25			T = 50			T = 100		
				α			α			α		
				0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
$\lambda = 0.5$	RMSE	$X_T^{(\mathrm{Mn})}(1)$	YW	0.89	0.96	1.31	0.90	0.95	1.14	0.90	0.96	0.91
		$X_T^{(\mathrm{Mn})}(1)$	CML	0.89	0.95	1.23	0.89	0.95	1.10	0.90	0.96	0.90
		$X_T^{(\mathrm{Mn})}(1)$	CLS	0.89	0.96	1.27	0.90	0.95	1.12	0.90	0.96	0.90
		$X_T^{(\mathrm{Md})}(1)$	YW	0.91	1.00	1.32	0.92	0.98	1.14	0.92	0.97	0.91
		$X_T^{(\mathrm{Md})}(1)$	CML	0.91	0.98	1.23	0.92	0.96	1.10	0.92	0.96	0.90
		$X_T^{(\mathrm{Md})}(1)$	CLS	0.92	1.00	1.27	0.92	0.98	1.12	0.92	0.97	0.90
		$X_T^{(\mathrm{Ei})}(1)$	YW	0.92	1.07	1.41	0.92	1.06	1.27	0.93	1.07	1.07
		$X_T^{(\mathrm{Ei})}(1)$	CML	0.92	1.05	1.37	0.92	1.05	1.25	0.93	1.07	1.06
		$X_T^{(\mathrm{Ei})}(1)$	CLS	0.92	1.06	1.38	0.92	1.06	1.25	0.93	1.07	1.06
	MAE	$X_T^{(\mathrm{Mn})}(1)$	YW	0.59	0.67	0.78	0.56	0.67	0.72	0.56	0.64	0.70
		$X_T^{(\mathrm{Mn})}(1)$	CML	0.59	0.66	0.68	0.56	0.66	0.68	0.56	0.64	0.67
		$X_T^{(\mathrm{Mn})}(1)$	CLS	0.59	0.66	0.72	0.56	0.67	0.69	0.56	0.64	0.68
		$X_T^{(\mathrm{Md})}(1)$	YW	0.57	0.68	0.78	0.55	0.68	0.71	0.55	0.64	0.69
		$X_T^{(\mathrm{Md})}(1)$	CML	0.57	0.66	0.68	0.55	0.67	0.67	0.55	0.63	0.67
		$X_T^{(\mathrm{Md})}(1)$	CLS	0.57	0.68	0.73	0.55	0.68	0.69	0.55	0.64	0.68
		$X_T^{(\mathrm{Ei})}(1)$	YW	0.57	0.73	0.88	0.55	0.74	0.84	0.55	0.71	0.82
		$X_T^{(\mathrm{Ei})}(1)$	CML	0.57	0.72	0.85	0.55	0.73	0.83	0.55	0.71	0.82
		$X_T^{(\mathrm{Ei})}(1)$	CLS	0.57	0.72	0.86	0.55	0.73	0.83	0.55	0.71	0.82
$\lambda = 1$	RMSE	$X_T^{(\mathrm{Mn})}(1)$	YW	1.11	1.31	1.80	1.07	1.30	1.59	1.06	1.28	1.36
		$X_T^{(\mathrm{Mn})}(1)$	CML	1.12	1.30	1.68	1.08	1.29	1.54	1.06	1.27	1.34
		$X_T^{(\mathrm{Mn})}(1)$	CLS	1.12	1.31	1.74	1.08	1.30	1.57	1.06	1.28	1.35
		$X_T^{(\mathrm{Md})}(1)$	YW	1.22	1.35	1.80	1.20	1.34	1.59	1.17	1.32	1.36
		$X_T^{(\mathrm{Md})}(1)$	CML	1.22	1.34	1.68	1.20	1.33	1.54	1.18	1.31	1.34
		$X_T^{(\mathrm{Md})}(1)$	CLS	1.23	1.35	1.74	1.20	1.34	1.57	1.18	1.32	1.35
		$X_T^{(\mathrm{Ei})}(1)$	YW	1.23	1.39	1.87	1.21	1.39	1.68	1.18	1.37	1.45
		$X_T^{(\mathrm{Ei})}(1)$	CML	1.24	1.38	1.79	1.22	1.38	1.63	1.19	1.37	1.44
		$X_T^{(\mathrm{Ei})}(1)$	CLS	1.24	1.38	1.81	1.21	1.38	1.64	1.18	1.37	1.44
	MAE	$X_T^{(\mathrm{Mn})}(1)$	YW	0.80	0.98	1.18	0.79	0.95	1.09	0.78	0.94	1.06
		$X_T^{(\mathrm{Mn})}(1)$	CML	0.81	0.98	1.06	0.79	0.95	1.04	0.78	0.94	1.04
		$X_T^{(\mathrm{Mn})}(1)$	CLS	0.81	0.99	1.12	0.79	0.95	1.07	0.78	0.94	1.05
		$X_T^{(\mathrm{Md})}(1)$	YW	0.85	1.00	1.18	0.86	0.97	1.09	0.84	0.96	1.05
		$X_T^{(\mathrm{Md})}(1)$	CML	0.86	0.99	1.06	0.86	0.97	1.04	0.84	0.95	1.04
		$X_T^{(\mathrm{Md})}(1)$	CLS	0.86	1.00	1.12	0.86	0.98	1.06	0.84	0.96	1.05
		$X_T^{(\mathrm{Ei})}(1)$	YW	0.86	1.03	1.25	0.87	0.99	1.16	0.84	0.99	1.12
		$X_T^{(\mathrm{Ei})}(1)$	CML	0.87	1.02	1.17	0.87	0.99	1.12	0.85	0.99	1.12
		$X_T^{(\mathrm{Ei})}(1)$	CLS	0.87	1.03	1.19	0.87	0.99	1.13	0.85	1.00	1.12

Table 3.2: RMSE and MAE of $X_T^{(\mathrm{Mn})}(1)$, $X_T^{(\mathrm{Md})}(1)$ and $X_T^{(\mathrm{Ei})}(1)$ for different values of α and $\lambda = 0.5, 1$, using three estimation methods, YW, CML and CLS, for sample sizes $T = 25$, 50 and 100.

	Error	Forecast	Estimator	$T=25$			$T=50$			$T=100$		
				α			α			α		
				0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
$\lambda=3$	RMSE	$X_T^{(\mathrm{Mn})}(1)$	YW	1.89	2.19	3.00	1.87	2.23	3.79	1.85	2.09	2.13
		$X_T^{(\mathrm{Mn})}(1)$	CML	1.90	2.18	2.82	1.87	2.21	3.74	1.85	2.08	2.10
		$X_T^{(\mathrm{Mn})}(1)$	CLS	1.90	2.20	2.89	1.87	2.23	3.76	1.85	2.09	2.12
		$X_T^{(\mathrm{Md})}(1)$	YW	1.94	2.22	3.00	1.92	2.25	3.79	1.89	2.11	2.13
		$X_T^{(\mathrm{Md})}(1)$	CML	1.95	2.20	2.82	1.92	2.24	3.74	1.89	2.11	2.10
		$X_T^{(\mathrm{Md})}(1)$	CLS	1.95	2.23	2.90	1.92	2.25	3.76	1.89	2.11	2.12
		$X_T^{(\mathrm{Ei})}(1)$	YW	1.94	2.25	3.04	1.92	2.28	3.82	1.89	2.14	2.19
		$X_T^{(\mathrm{Ei})}(1)$	CML	1.96	2.23	2.88	1.93	2.27	3.77	1.89	2.14	2.15
		$X_T^{(\mathrm{Ei})}(1)$	CLS	1.95	2.25	2.94	1.92	2.28	3.79	1.89	2.14	2.17
	MAE	$X_T^{(\mathrm{Mn})}(1)$	YW	1.47	1.74	2.09	1.42	1.72	1.96	1.43	1.69	1.92
		$X_T^{(\mathrm{Mn})}(1)$	CML	1.49	1.74	1.93	1.42	1.70	1.90	1.44	1.69	1.90
		$X_T^{(\mathrm{Mn})}(1)$	CLS	1.48	1.75	2.00	1.42	1.72	1.93	1.43	1.69	1.90
		$X_T^{(\mathrm{Md})}(1)$	YW	1.49	1.75	2.10	1.44	1.72	1.96	1.45	1.70	1.91
		$X_T^{(\mathrm{Md})}(1)$	CML	1.50	1.75	1.93	1.45	1.71	1.90	1.46	1.70	1.90
		$X_T^{(\mathrm{Md})}(1)$	CLS	1.50	1.77	2.00	1.45	1.72	1.93	1.45	1.70	1.90
		$X_T^{(\mathrm{Ei})}(1)$	YW	1.49	1.76	2.13	1.44	1.74	2.00	1.46	1.72	1.95
		$X_T^{(\mathrm{Ei})}(1)$	CML	1.50	1.76	1.99	1.45	1.73	1.93	1.46	1.71	1.92
		$X_T^{(\mathrm{Ei})}(1)$	CLS	1.50	1.77	2.04	1.45	1.74	1.95	1.46	1.72	1.93
$\lambda=5$	RMSE	$X_T^{(\mathrm{Mn})}(1)$	YW	2.44	2.92	4.16	2.40	2.80	3.80	2.39	2.72	2.28
		$X_T^{(\mathrm{Mn})}(1)$	CML	2.46	2.90	3.93	2.41	2.79	3.71	2.39	2.71	2.23
		$X_T^{(\mathrm{Mn})}(1)$	CLS	2.46	2.92	4.03	2.40	2.80	3.75	2.39	2.72	2.26
		$X_T^{(\mathrm{Md})}(1)$	YW	2.48	2.94	4.16	2.44	2.82	3.80	2.42	2.74	2.28
		$X_T^{(\mathrm{Md})}(1)$	CML	2.50	2.92	3.93	2.44	2.81	3.71	2.42	2.73	2.23
		$X_T^{(\mathrm{Md})}(1)$	CLS	2.50	2.94	4.03	2.44	2.82	3.75	2.42	2.74	2.26
		$X_T^{(\mathrm{Ei})}(1)$	YW	2.48	2.96	4.19	2.44	2.85	3.83	2.42	2.76	2.34
		$X_T^{(\mathrm{Ei})}(1)$	CML	2.50	2.93	3.97	2.45	2.83	3.74	2.42	2.75	2.28
		$X_T^{(\mathrm{Ei})}(1)$	CLS	2.50	2.96	4.06	2.44	2.84	3.78	2.42	2.76	2.31
	MAE	$X_T^{(\mathrm{Mn})}(1)$	YW	1.92	2.28	2.69	1.89	2.19	2.60	1.87	2.13	2.48
		$X_T^{(\mathrm{Mn})}(1)$	CML	1.94	2.25	2.46	1.89	2.18	2.50	1.87	2.13	2.45
		$X_T^{(\mathrm{Mn})}(1)$	CLS	1.93	2.28	2.55	1.89	2.18	2.55	1.87	2.13	2.47
		$X_T^{(\mathrm{Md})}(1)$	YW	1.94	2.29	2.69	1.90	2.19	2.60	1.87	2.14	2.48
		$X_T^{(\mathrm{Md})}(1)$	CML	1.96	2.27	2.46	1.90	2.18	2.50	1.87	2.14	2.45
		$X_T^{(\mathrm{Md})}(1)$	CLS	1.95	2.30	2.56	1.89	2.19	2.55	1.87	2.14	2.47
		$X_T^{(\mathrm{Ei})}(1)$	YW	1.94	2.30	2.71	1.90	2.21	2.62	1.87	2.16	2.52
		$X_T^{(\mathrm{Ei})}(1)$	CML	1.96	2.28	2.50	1.90	2.19	2.52	1.88	2.15	2.48
		$X_T^{(\mathrm{Ei})}(1)$	CLS	1.95	2.31	2.58	1.90	2.20	2.57	1.88	2.16	2.50

Table 3.3: RMSE and MAE of $X_T^{(\mathrm{Mn})}(1)$, $X_T^{(\mathrm{Md})}(1)$ and $X_T^{(\mathrm{Ei})}(1)$ for different values of α and $\lambda=3,5$, using three estimation methods, YW, CML and CLS, for sample sizes $T=20$, 50 and 100.

3.3 Monte Carlo results for perturbation study

In order to study the predictive power of Poisson INAR(1) model under misspecified data, we simulated 10000 series of lengths $T = 25, 50, 100, 200, 300$ and 500 using binomial thinning with parameters $\alpha = 0.1, 0.5$ and 0.9 and misspecifying the arrival process $(\epsilon_t)_{t \geqslant 1}$ by letting its distribution be uniform over $\{0, 1, 2, 3\}$. So, the distribution of ϵ_t does not depend on unknown parameters. Freeland [1998] states that for the misspecified model, the parameter λ can viewed as the mean parameter of ϵ_t, however it is not used to specify the distribution of ϵ_t, thus, estimates of λ are estimates of the mean of ϵ_t, this is $3/2$.

We calculated the RMSE and the MAE of conditional median using the true model and the misspecified model. The true model corresponds to the case when we estimate the parameter α taking into account that ϵ_t has uniform distribution over $\{0, 1, 2, 3\}$; in this case, the CLS and CML estimators of α are different of the CLS and CML estimators of α given in Section 2.3, while the YW estimator of α does not change. For the true model we found $X_T^{(\mathrm{Mn})}(1)$, the conditional median of the distribution given by the convolution of a binomial distribution with parameters α and X_T with the uniform distribution over $\{0, 1, 2, 3\}$. The misspecified model corresponds to the case when we estimate α and λ assuming that the series satisfies the Poisson INAR(1) model.

Table 3.4 shows the RMSE and the MAE of conditional median for the true and misspecified models. $\mathrm{PM}_{\mathrm{YW}}$ denotes the misspecified model with conditional distribution given by the convolution of the binomial distribution with parameters $\widehat{\alpha}_{\mathrm{YW}}$ and X_T with the Poisson distribution with parameter $\widehat{\lambda}_{\mathrm{YW}}$, and $\mathrm{UM}_{\mathrm{YW}}$ denotes the true model with conditional distribution given by the convolution of the binomial distribution with parameters $\widehat{\alpha}_{\mathrm{YW}}$ and X_T with the uniform distribution over $\{0, 1, 2, 3\}$. Similarly, PM_{CLS} denotes the misspecified model with conditional distribution given by the convolution of the binomial distribution with parameters $\widehat{\alpha}_{\mathrm{CLS}}$ and X_T with a Poisson distribution with parameter $\widehat{\lambda}_{\mathrm{CLS}}$, and UM_{CLS} denotes the true model with conditional distribution given by the convolution of the binomial distribution with parameters $\widehat{\alpha}_{\mathrm{CML}}$ and X_T with the uniform distribution over $\{0, 1, 2, 3\}$. Analogously, PM_{CML} denotes the misspecified model with conditional distribution given by the convolution of the binomial distribution with parameters $\widehat{\alpha}_{\mathrm{CML}}$ and X_T with the Poisson distribution with parameter $\widehat{\lambda}_{\mathrm{CML}}$, and UM_{CML} denotes the true model with conditional distribution given by the convolution of the binomial distribution with parameters $\widehat{\alpha}_{\mathrm{CML}}$ and X_T with the uniform distribution over $\{0, 1, 2, 3\}$. Here $\widehat{\alpha}_{\mathrm{YW}}$, $\widehat{\alpha}_{\mathrm{CLS}}$, $\widehat{\alpha}_{\mathrm{CML}}$ are the estimates of α obtained from YW, CLS and CML methods, respectively, and $\widehat{\lambda}_{\mathrm{YW}}$, $\widehat{\lambda}_{\mathrm{CLS}}$, $\widehat{\lambda}_{\mathrm{CML}}$ are the estimates of λ using YW, CLS and CML methods.

Note that for CLS and CML estimators, the RMSE and the MAE of true model are very close to RMSE and MAE of the misspecified model, respectively. The same happens for YW estimators when $\alpha = 0.1$, while for $\alpha = 0.5$ and $\alpha = 0.9$ the RMSE and the MAE of the misspecified model are smaller than the RMSE and MAE of the true model for small values of T and the RMSE and MAE are close as T increases. This means that the predictive power does not deteriorate even when the misspecified model is used.

Error	Estimator	$T=25$			$T=50$			$T=100$			$T=200$			$T=300$			$T=500$		
		α			α			α			α			α			α		
		0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
RMSE	PM_{YW}	1.28	1.51	1.79	1.26	1.48	1.70	1.29	1.47	1.65	1.27	1.46	1.66	1.28	1.46	1.62	1.26	1.47	1.62
	PM_{CLS}	1.28	1.51	1.72	1.26	1.48	1.69	1.29	1.47	1.64	1.27	1.46	1.66	1.28	1.46	1.62	1.26	1.47	1.62
	PM_{CML}	1.29	1.50	1.66	1.26	1.47	1.66	1.29	1.47	1.64	1.26	1.47	1.65	1.28	1.46	1.63	1.26	1.47	1.63
	UM_{YW}	1.27	1.58	4.30	1.23	1.51	2.66	1.25	1.49	1.97	1.24	1.47	1.78	1.24	1.46	1.69	1.23	1.47	1.66
	UM_{CLS}	1.25	1.48	1.66	1.22	1.46	1.66	1.25	1.47	1.64	1.24	1.46	1.66	1.24	1.45	1.63	1.23	1.47	1.62
	UM_{CML}	1.29	1.49	1.66	1.23	1.46	1.65	1.25	1.47	1.64	1.24	1.46	1.66	1.24	1.45	1.63	1.23	1.47	1.62
MAE	PM_{YW}	1.03	1.20	1.40	1.02	1.17	1.33	1.05	1.16	1.29	1.03	1.18	1.31	1.03	1.16	1.27	1.02	1.16	1.26
	PM_{CLS}	1.04	1.20	1.35	1.01	1.18	1.32	1.06	1.16	1.29	1.03	1.18	1.30	1.03	1.16	1.27	1.02	1.16	1.26
	PM_{CML}	1.04	1.19	1.30	1.02	1.17	1.30	1.06	1.16	1.28	1.02	1.18	1.29	1.03	1.16	1.28	1.03	1.16	1.27
	UM_{YW}	1.03	1.24	3.43	1.00	1.20	2.08	1.03	1.18	1.55	1.01	1.18	1.39	1.01	1.16	1.32	1.01	1.17	1.29
	UM_{CLS}	1.02	1.16	1.29	1.00	1.16	1.29	1.03	1.16	1.29	1.01	1.18	1.29	1.01	1.15	1.27	1.01	1.17	1.26
	UM_{CML}	1.04	1.17	1.30	1.00	1.16	1.29	1.03	1.16	1.29	1.01	1.17	1.30	1.01	1.15	1.28	1.01	1.17	1.26

Table 3.4: RMSE and MAE of $X_T^{(\mathrm{Md})}(1)$ for different values of α using YW, CLS and CML estimators considering uniform distribution of arrival process, for sample sizes $T = 25, 50, 100, 200, 300$ and 500.

3.4 Applications

We apply the presented methodology to two data sets, both of which having been obtained from the Workers Compensation Board (WCB) of British Columbia, Canada. The WCB provides disability insurance for more than 130,000 employers in British Columbia. Every year the WCB receives about 200,000 new claims. These data sets were already investigated by Freeland [1998]. Each of the two series here studied contains 120 monthly counts of claimants collecting Short Term Wage Loss Benefits from the WCB from January 1985 to December 1994. All the claimants are male, between the ages of 35 and 54, work in the logging industry and reported their claim to the Richmond service delivery location. The distinguishing difference between the two series is the nature of the injury. The first data set relates to claimants who have had soft tissue injures, such as contusions and bruises, while the second data set relates to claimants with dislocations. We refer to the two data sets as SOFT INJURES and DISLOCATIONS respectively.

Clearly these data may be considered as INAR(1) processes. That is, at any month t, the observed number of claimants, X_t, can be viewed as the sum of the number of claimants from the previous period surviving in the claims queue, $\alpha \circ X_{t-1}$, and the number of newly injured workers ϵ_t.

Data	Minimum Count	Maximum Count	Median	Mean	Variance
SOFT INJURES	4	17	9	9.825	9.478
DISLOCATIONS	0	4	1	0.917	0.760

Table 3.5: Summary Statistics for SOFT INJURES and DISLOCATIONS data

Figure 3.1 and Figure 3.2 provide the time series plots of SOFT INJURES and DISLOCATIONS data as well as their corresponding sample autocorrelation and sample partial autocorrelation functions. A summary of their simple descriptive statistics is reported in Table 3.5. Note that both data sets have mean near variance; for the SOFT INJURES data the sample partial autocorrelation function is consistent with the INAR(1) model and, for the DISLOCATIONS data, from the sample correlation and partial autocorrelation functions, it becomes clear that a first order autoregressive model seems to be reasonable.

In order to compare the different forecasts, we found the one step ahead point prediction of monthly claims count from May 1993 to December 1994. Table 3.6 and Table 3.7 present the point forecasts based on the mean, median and mode of the conditional distribution of X_{T+1} given X_T as well as the CLS and CML estimates for α and λ for SOFT INJURES and DISLOCATIONS data, respectively. For SOFT INJURES data, note that the MSE and MAE of the integer part of conditional mean $x_T^{(\mathrm{Ei})}(1)$ are a little less than the MSE and MAE of conditional median and conditional mode, i. e., the integer part of conditional mean is a little better than the other forecasts. For DISLOCATIONS data, note that, using CLS estimates, the conditional median is better than conditional mode and integer part of conditional mean, while using CML estimates the conditional median and mode are competitive and they are better than the integer part of conditional mean.
Furthermore, note that for the SOFT INJURES data

$$\text{MSE of } x_T^{(\mathrm{Ei})}(1) \text{ / MSE of } x_T^{(\mathrm{Md})}(1) = 4.85/5.2 \approx 0.93$$

and for the DISLOCATIONS data

$$\text{MSE of } x_T^{(\mathrm{Md})}(1) \text{ / MSE of } x_T^{(\mathrm{Ei})}(1) = 0.4/0.75 \approx 0.53,$$

so, in the data set where the conditional median and mode are better they produce a reduction of 46% of the MSE, while in the data set where the integer part of the conditional mean is better it produces only a reduction of 7% of the MSE. Then, it can be concluded that the conditional median is better than the other two forecasts.

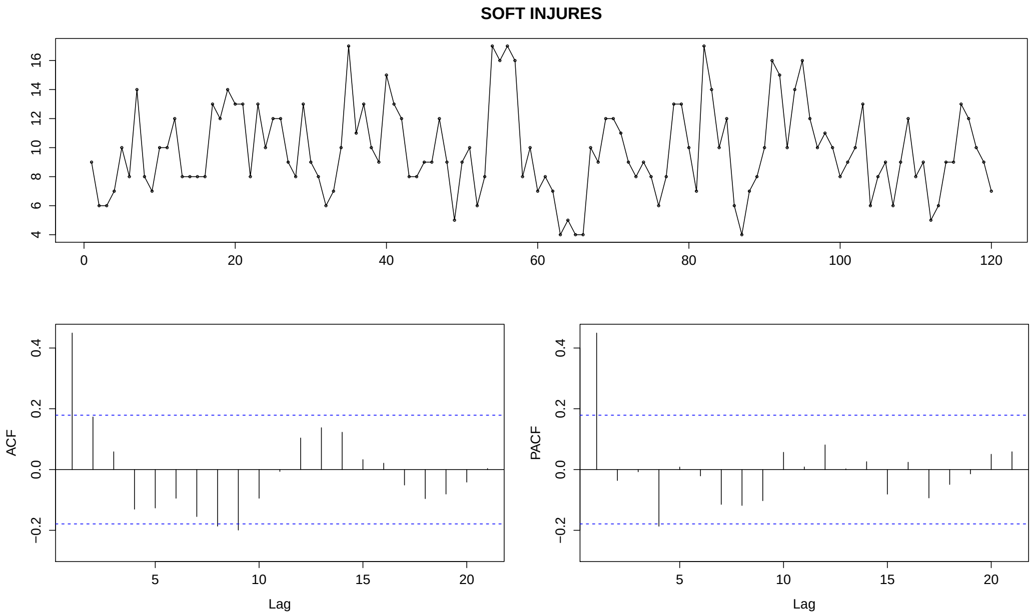

Figure 3.1: Monthly counts of SOFT INJURES data, January 1985-December 1994 and sample autocorrelation and partial autocorrelation functions.

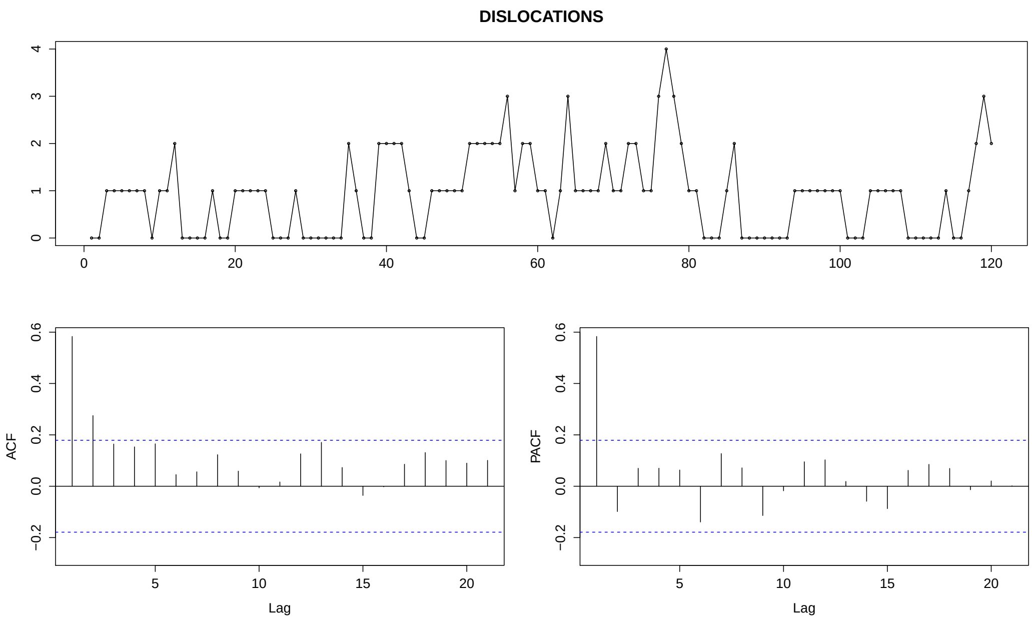

Figure 3.2: Monthly counts of DISLOCATIONS data, January 1985-December 1994 and sample autocorrelation and partial autocorrelation functions.

		CLS		CML			CLS			CML		
Month/Year	T	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\lambda}$	x_{T+1}	$x_T^{(Md)}(1)$	$x_T^{(Md)}(1)$	$x_T^{(Ei)}(1)$	$x_T^{(Md)}(1)$	$x_T^{(Md)}(1)$	$x_T^{(Ei)}(1)$
May/93	100	0.467	5.326	0.473	5.271	9	9	9	9	9	9	9
Jun/93	101	0.468	5.324	0.476	5.241	10	9	9	9	9	9	9
Jul/93	102	0.467	5.333	0.479	5.218	13	10	10	10	10	10	10
Aug/93	103	0.467	5.362	0.480	5.238	6	11	11	11	11	11	11
Set/93	104	0.451	5.468	0.459	5.394	8	8	8	8	8	8	8
Oct/93	105	0.452	5.459	0.462	5.365	9	9	9	9	9	9	9
Nov/93	106	0.452	5.457	0.464	5.336	6	9	9	9	9	9	9
Dec/93	107	0.455	5.392	0.463	5.316	9	8	8	8	8	8	8
Jan/94	108	0.452	5.432	0.463	5.327	12	9	9	9	9	9	9
Feb/94	109	0.450	5.476	0.463	5.350	8	11	11	10	11	11	10
Mar/94	110	0.444	5.505	0.458	5.372	9	9	9	9	9	9	9
Apr/94	111	0.445	5.504	0.461	5.344	5	9	9	9	9	9	9
May/94	112	0.448	5.426	0.456	5.349	6	8	7	7	7	7	7
Jun/94	113	0.456	5.338	0.463	5.271	9	8	8	8	8	8	8
Jul/94	114	0.453	5.377	0.462	5.282	9	9	9	9	9	9	9
Aug/94	115	0.453	5.370	0.465	5.251	13	9	9	9	9	9	9
Set/94	116	0.450	5.426	0.463	5.302	12	11	11	11	11	11	11
Oct/94	117	0.452	5.413	0.467	5.266	10	11	11	10	11	11	10
Nov/94	118	0.451	5.421	0.469	5.245	9	10	10	9	10	10	9
Dec/94	119	0.451	5.415	0.471	5.217	7	9	9	9	9	9	9
-	-	-	-	-	-	MSE	5.35	5.2	4.85	5.2	5.2	4.85
-	-	-	-	-	-	MAE	1.75	1.7	1.55	1.7	1.7	1.55

Table 3.6: Point prediction one step ahead of monthly claims count May 1993 to December 1994 for SOFT INJURES data.

		CLS		CML			CLS			CML		
Month/Year	T	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\lambda}$	x_{T+1}	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Ei})}(1)$	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Ei})}(1)$
May/93	100	0.565	0.427	0.657	0.339	0	1	1	0	1	1	0
Jun/93	101	0.565	0.418	0.647	0.339	0	0	0	0	0	0	0
Jul/93	102	0.570	0.409	0.649	0.334	0	0	0	0	0	0	0
Aug/93	103	0.575	0.400	0.650	0.329	1	0	0	0	0	0	0
Set/93	104	0.568	0.412	0.647	0.339	1	1	1	0	1	1	0
Oct/93	105	0.568	0.413	0.651	0.335	1	1	1	0	1	1	0
Nov/93	106	0.568	0.413	0.656	0.331	1	1	1	0	1	1	0
Dez/93	107	0.568	0.413	0.660	0.327	1	1	1	0	1	1	0
Jan/94	108	0.568	0.413	0.664	0.323	0	1	1	0	1	1	0
Fev/94	109	0.567	0.405	0.655	0.323	0	0	0	0	0	0	0
Mar/94	110	0.572	0.397	0.656	0.319	0	0	0	0	0	0	0
Apr/94	111	0.576	0.389	0.658	0.314	0	0	0	0	0	0	0
May/94	112	0.581	0.381	0.659	0.310	0	0	0	0	0	0	0
Jun/94	113	0.585	0.374	0.660	0.306	1	0	0	0	0	0	0
Jul/94	114	0.578	0.386	0.657	0.315	0	1	1	0	1	1	0
Aug/94	115	0.577	0.378	0.648	0.315	0	0	0	0	0	0	0
Set/94	116	0.581	0.372	0.649	0.311	1	0	0	0	0	0	0
Oct/94	117	0.574	0.383	0.646	0.320	2	1	1	0	1	1	0
Nov/94	118	0.576	0.390	0.647	0.327	3	2	1	1	2	2	1
Dez/94	119	0.595	0.386	0.653	0.334	2	2	2	2	2	2	2
-	-	-	-	-	-	MSE	0.4	0.55	0.75	0.4	0.4	0.75
-	-	-	-	-	-	MAE	0.4	0.45	0.55	0.4	0.4	0.55

Table 3.7: Point prediction one step ahead of monthly claims count May 1993 to December 1994 for DISLOCATIONS data.

Chapter 4

Forecasting for INARCH(1)

Resumo

Começamos este capítulo apresentando o operador thinning Poisson junto com suas propriedades. Na seção 4.2, apresentamos uma forma alternativa de definir o processo INARCH(1), baseado no operador thinning Poisson. Na seção 4.3 fornecemos uma prova da existência e unicidade da distribuição estacionaria marginal do processo INARCH(1), apresentamos expressões analíticas para o $r-$ésimo momento ordinário marginal, para a média e a variância condicional e para a função geradora de probabilidades $h-$passos à frente. Além disso, para o caso particular $h = 2$, encontramos uma expressão simples para a função de probabilidade condicional dois passos à frente. Dado que a distribuição condicional um passo à frente do processo INARCH(1) é uma distribuição Poisson, nós propomos sua média e sua moda como previsões um passo à frente na seção 4.4. Embora a moda de uma distribuição Poisson tenha uma expressão analítica simples, não existe uma expressão para sua mediana. Portanto, usando estudos de simulação não apresentados aqui, nós propomos uma aproximação simples da mediana da distribuição Poisson, que tem um bom desempenho, em termos de erro quadrático médio e de erro absoluto médio. Na seção 4.3 apresentamos propriedades tais como média e variância limites da mediana aproximada. Começamos provando consistência condicional fraca da mediana aproximada estimada e então

conseguimos provar sua consistência condicional forte. Também, provamos consistência fraca da mediana aproximada estimada e então conseguimos provar sua consistência forte. Embora não tenhamos uma expressão analítica viável para a função de probabilidade condicional $h-$passos à frente para $h \geqslant 3$, na seção 4.5 propomos uma forma recursiva de obter a previsão $h-$passos à frente para $h \geqslant 2$. Na seção 4.6 apresentamos uma distribuição que nos permite obter intervalos de previsão unilaterais e bilaterais. Nas seções 4.7, 4.8 e 4.9 apresentamos simulações de Monte Carlo que comparam os desempenhos das previsões propostas. No final do capítulo ilustramos as metodologias propostas usando dois conjuntos de dados reais que já têm sido considerados neste processo.

Initial presentation

We begin this chapter defining the Poisson thinning operator and then we find and prove its properties. In Section 4.2, we present an alternative way to define the INARCH(1)process, based on the Poisson thinning operator. In Section 4.3 we provide a proof of existence and uniqueness of the marginal stationary distribution of the INARCH(1) process, we obtain analytic expressions for the $r-$th marginal ordinary moment, for the $h-$steps conditional mean and variance as well as for the $h-$steps ahead conditional probability generating function. Besides, for the particular case $h = 2$, we find a simple expression for the two-steps ahead conditional probability function. Given that the INARCH(1) process has the advantage that the conditional distribution one-step ahead is a Poisson distribution, we propose its median and mode as forecasts one-step ahead in Section 4.4. Although the mode of a Poisson distribution has an easy analytic expression, there is no expression for the median. Hence, by simulation study, which does not present in this work, we propose an easy approximation of the median of a Poisson distribution which works very well in terms of mean squared error and mean absolute error. In Section 4.3 we show properties such as mean and variance limits of the approximate median. We begin with the proof of weakly conditional consistency of the approximate median, and then we get to prove strongly conditional consistency. Further, we are able to demonstrate weakly consistency of the approximate median and then we get to prove its strongly consistency.Although we obtain an analytical expression for the $h-$steps ahead conditional probability generating function in Section 4.3, it does not lead to a workable procedure to obtain the $h-$steps ahead conditional probability function for $h \geqslant 3$, so, in Section 4.5 we propose a recursive way to find the $h-$steps ahead forecast for $h \geqslant 2$. In Section 4.6 we show a distribution which allows to obtain one-sided and two-sided predictions intervals. In Sections 4.7, 4.8 and 4.9 we present Monte Carlo simulation studies that compare the behaviors of the proposed forecasts. At the end of this chapter we illustrate the proposed approaches with two different real data sets, which were studied in this process.

4.1 The Poisson thinning operator

Definition 5. *Let X be a non-negative integer-valued random variable and $\alpha \geqslant 0$. We define the Poisson thinning operator as*

$$\alpha * X \stackrel{d}{=} \sum_{i=1}^{X} N_i, \tag{4.1}$$

where the N_i are i.i.d. Poisson random variables, independent of X, with parameter α, and the notation $X \stackrel{d}{=} Y$ means that X has the same distribution as Y.

The sequence $N_1, N_2, \ldots$ is said to be the counting series of $\alpha * X$. From definition above it is clear that

$$\alpha * X \,|\, X \sim \text{ Po}(\alpha X).$$

The difference between this definition and Definition 1 is the probability distribution of the counting series. In Definition 1 the counting series is Bernoulli distributed, thus its parameter belongs to interval [0, 1], however in Definition 4.1 the counting series has Poisson distribution, so its parameter belongs to interval $(0, \infty)$.

So, we call the '$*$' operator *the Poisson thinning operator* and it can be interpreted as follows: consider a population of size X at a certain time t. Suppose that in this population each individual dies or produces independently offsprings according to the

Poisson distribution with parameter α. Let N_i defined as

$$N_i = \begin{cases} 0, & \text{if the individual } i \text{ died,} \\ 1, & \text{if the individual } i \text{ had no offsprings,} \\ m, & \text{if the individual } i \text{ had } m-1 \text{ offsprings, for } m \geqslant 2; \end{cases}$$

then, if we observe the same population in the later time $t+1$, the population size in time $t+1$ is given by $\alpha * X$.

The properties of the Poisson thinning operator are presented in the next lemma.

Lemma 2. *Let X_i $(i = 1, \dots, m)$ be a sequence of non-negative integer-valued identically distributed random variables, α_i $(i = 1, \dots, m)$ a sequence of non-negative real constants and suppose that the counting series of $\alpha_i * X_i$ are mutually independent, identically distributed according to* $\mathrm{Po}(\alpha_i)$*, and independent of X_i. Then,*

i) $0 * X_1 = 0$

ii) $\alpha_1 * \left(X_1 + X_2\right) \overset{d}{=} \alpha_1 * X_1 + \alpha_1 * X_2$ *if X_1 and X_2 are independent*

iii) $\mathrm{E}\left[\alpha_1 * X_1\right] = \alpha_1 \mathrm{E}\left[X_1\right]$

iv) $\mathrm{Var}\left[\alpha_1 * X_1\right] = \alpha_1^2 \mathrm{Var}\left[X_1\right] + \alpha_1 \mathrm{E}\left[X_1\right]$

v) $\mathrm{E}\left[\alpha_1 * X_1 \,|\, X_1\right] = \alpha_1 X_1$

vi) $\mathrm{Var}\left[\alpha_1 * X_1 \,|\, X_1\right] = \alpha_1 X_1$

vii) $\mathrm{Cov}\left(\alpha_1 * X_1, \alpha_2 * X_2\right) = \alpha_1 \alpha_2 \mathrm{Cov}\left(X_1, X_2\right)$

viii) $\mathrm{E}\left[\left(\alpha_1 * X_1\right)^2\right] = \alpha_1 \mathrm{E}\left[X_1\right] + \alpha_1^2 \mathrm{E}\left[X_1^2\right]$

ix) $\mathrm{E}\left[\left(\alpha_1 * X_1\right)^r\right] = \sum_{k=1}^{r} S(r,k)\, \alpha_1^k \mathrm{E}\left[X_1^k\right]$, *where* $S(r,k) = 1/k! \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^r$ *is the Stirling number of the second kind, i.e., the number of ways to partition a set of r objects into k non-empty subsets.*

x) $\mathrm{E}\left[(\alpha_1 * X_1) X_2\right] = \alpha_1 \mathrm{E}\left[X_1 X_2\right]$

xi) $\mathrm{E}\left[(\alpha_1 * X_1)^2 X_2\right] = \alpha_1 \mathrm{E}\left[X_1 X_2\right] + \alpha_1^2 \mathrm{E}\left[X_1^2 X_2\right]$

xii) $\mathrm{E}\big[(\alpha_1 * X_1)^r\, X_2\,\big] \;=\; \sum_{k=1}^{r}\, S(r,k)\,\alpha_1^k\,\mathrm{E}\big[\,X_1^k\, X_2\,\big]$

xiii) $\mathrm{E}\big[(\alpha_1 * X_1)\;\prod_{i=2}^{m} X_i\,\big] \;=\; \alpha_1\,\mathrm{E}\big[\,\prod_{i=1}^{m} X_i\,\big]$

xiv) $\mathrm{E}\big[(\alpha_1 * X_1)^r\;\prod_{i=2}^{m} X_i\,\big] \;=\; \sum_{k=1}^{r}\; S(r,k)\,\alpha_1^k\,\mathrm{E}\big[X_1^k\;\prod_{i=1}^{m} X_i\,\big]$

xv) $\mathrm{E}\big[(\alpha_1 * X_1\,)(\alpha_2 * X_2\,)\big] \;=\; \alpha_1\,\alpha_2\,\mathrm{E}\,\big[\,X_1\,X_2\big]$

xvi) $\mathrm{E}\big[\,\prod_{i=1}^{m}\,(\alpha_i * X_i)\,\big] \;=\; \big(\prod_{i=1}^{m}\alpha_i\big)\,\mathrm{E}\,\big[\,\prod_{i=1}^{m} X_i\big]$

xvii) $\mathrm{E}\big[(\alpha_1 * X_1\,)^2(\alpha_2 * X_2\,)\big] \;=\; \alpha_1^2\,\alpha_2\,\mathrm{E}\,\big[\,X_1^2\,X_2\big] + \alpha_1\,\alpha_2\,\mathrm{E}\,\big[\,X_1\,X_2\,\big]$

xviii) $\mathrm{E}\big[(\alpha_1 * X_1\,)^r(\alpha_2 * X_2\,)\big] \;=\; \alpha_2\,\sum_{k=1}^{r}\; S(r,k)\,\alpha_1^k\,\mathrm{E}\big[\,X_1^k\,X_2\,\big]$

xix) $\mathrm{E}\big[\alpha_1 * (\alpha_1 * X_1)\,\big] \;=\; \alpha_1^2\mathrm{E}\big[\,X_1\,\big]$

xx) $\mathrm{E}\big[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * X_1))}_{r\;\;\alpha_1\text{'s}}\,\big] \;=\; \alpha_1^r\,\mathrm{E}\big[\,X_1\,\big]$

xxi) $\mathrm{Var}\big[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * X_1))}_{r\;\;\alpha_1\text{'s}}\,\big] \;=\; \alpha_1^{2r}\,\mathrm{Var}\big[\,X_1\,\big] + \left(\dfrac{1-\alpha_1^r}{1-\alpha_1}\right)\alpha_1^r\,\mathrm{E}\big[X_1\big]$

xxii) $\mathrm{E}\big[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * X_1))}_{r\;\;\alpha_1\text{'s}}\;|\,X_1\,\big] \;=\; \alpha_1^r\,X_1$

xxiii) $\mathrm{Var}\big[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * X_1))}_{r\;\;\alpha_1\text{'s}}\;|\,X_1\,\big] \;=\; \alpha_1^r\,X_1\left(\dfrac{1-\alpha_1^r}{1-\alpha_1}\right)\;\text{for } r \geqslant 2$

The proofs of properties in Lemma 2 are given in Appendix. The properties *iv), v), vi), ix), xii), xviii), xix), xx), xxi), xxii)* and *xxiii)* are true for counting series with Poisson distribution and can be not true for counting series distributed with another discrete distribution.

4.2 INARCH(1) formulation model

Now we present the formulation of the INARCH(1) process presented in Section 2.5 based on the Poisson thinning operator defined above.

Definition 6. *A discrete non-negative integer-valued process $(X_t)_{t\geqslant 1}$, $X_t \in \mathbb{N}_0$, is said to be an INARCH(1) process if it satisfies the recursion*

$$X_t = \alpha * X_{t-1} + \epsilon_t \quad \textit{for} \;\; t \geqslant 1, \tag{4.2}$$

where $\alpha \geqslant 0$, '$*$' represents the Poisson thinning operator given in Definition 5, $(\epsilon_t)_{t\geqslant 1}$, $\epsilon_t \in \mathbb{N}_0$, is a sequence of of i.i.d. Poisson random variables with parameter λ and it is assumed that the counting series of $\alpha * X_{t-1}$ is independent of ϵ_t.

From property *i)* of Lemma 2 note that for $\alpha = 0$ the process is the sequence $(\epsilon_t)_{t\geqslant 1}$ and in this case the model is not overdispersed. On the other hand, in Section 2.5 was stated that the INARCH(1) process is a stationary processes whenever $\alpha < 1$. So, we should assume $\alpha \in (0, 1)$ to guarantee stationarity and overdispersion of the INARCH(1) process as it was assumed in Section 2.5.

The conditional probability distribution of X_t given X_{t-1} is Poisson with parameter $\alpha X_{t-1} + \lambda$ and the model can be interpreted as follow

$$\underbrace{X_t}_{\text{Population at time } t} = \underbrace{\alpha * X_{t-1}}_{\text{Population of time } t-1} + \underbrace{\epsilon_t}_{\text{Immigrations}} .$$

The following proposition asserts the existence and uniqueness of the marginal distribution of the INARCH(1) process.

Proposition 3. *Let $(X_t)_{t\geqslant 1}$, $X_t \in \mathbb{N}_0$, be the INARCH(1) process defined by recursion Equation* (4.2)*. Then, the marginal stationary distribution of X_t exists and it is unique.*

Proof. The INARCH(1) process can be seen as a Markov chain with transition probabilities π_{ij} given by

$$\pi_{ij} = \mathrm{P}(X_t = j \mid X_{t-1} = i) = \frac{(\alpha i + \lambda)^j \exp(-\alpha i - \lambda)}{j!}. \tag{4.3}$$

Our aim is to use [Ross [1983], Theorem 4.3.3 on p. 109] to prove the existence and uniqueness of the marginal stationary distribution of the INARCH(1) process.

We begin remembering that a Markov chain is said to be irreducible if all states communicate with each other; state i communicates with state j if and only if

$$\pi_{ij}^{n} > 0 \quad \text{and} \quad \pi_{ji}^{m} > 0 \quad \text{for some} \quad n, m \geqslant 0,$$

where $\pi_{ij}^{n} = \mathrm{P}(X_n = j \,|\, X_0 = i)$. Also, state i has period d if $d = \gcd\{n : \pi_{ii}^{n} > 0\}$, where gcd denotes the greatest common divisor. For $d = 1$, the Markov chain is said to be aperiodic.

Equation (4.3) shows that $\pi_{ij} > 0$ for all i, j, so $\pi_{ii} > 0$, which implies $d = 1$. Then, we have proved that $\left(X_t\right)_{t\geqslant 1}$ is an irreducible aperiodic Markov chain.

Using [Ross [1983], Theorem 4.3.3 on p. 109] if we get to prove that some state j is positive recurrent, then we can conclude that a stationary distribution for the INARCH(1) process exists and it is unique.

Now we will focus on proving that state 0 is a positive recurrent state; first we show that 0 is a recurrent state, this is equivalent to prove

$$\sum_{t=1}^{\infty} \pi_{00}^{t} = \infty, \quad \text{where} \quad \pi_{ij}^{t} = \mathrm{P}\big(X_{t+1} = j \,|\, X_1 = i\big).$$

We will prove by induction that the $(n+1)-$steps transition probability π_{k0}^{n+1} can be expressed as

$$\pi_{k0}^{n+1} = \exp(a_{n+1} + b_{n+1}k),$$

where $a_{n+1} = a_n + \lambda[\exp(b_n) - 1], b_{n+1} = \alpha[\exp(b_n) - 1], a_1 = -\lambda \quad \text{and} \quad b_1 = -\alpha.$

- ⋆ For $n = 0$, we have from equation (4.3) $\pi_{k0} = \exp(-\lambda - \alpha k)$, thus $a_1 = -\lambda$ and $b_1 = -\alpha$.
- ⋆ Suppose that $\pi_{k0}^{n} = \exp(a_n + b_n k)$
- ⋆ We can write the $(n+1)-$steps transition probability π_{k0}^{n+1} as

$$\begin{aligned}\pi_{k0}^{n+1} &= \sum_{i=0}^{\infty} \pi_{ki}\pi_{i0}^{n} \\ &= \sum_{i=0}^{\infty} \frac{\exp[-(\alpha k + \lambda)](\alpha k + \lambda)^{i}}{i!} \exp(a_n + b_n i)\end{aligned}$$

$$
\begin{aligned}
&= \exp[a_n - (\alpha k + \lambda)] \sum_{i=0}^{\infty} \frac{[(\alpha k + \lambda)\exp(b_n)]^i}{i!} \\
&= \exp\{a_n + \lambda[(\exp(b_n) - 1] + k\alpha[\exp(b_n) - 1]\} \\
&= \exp(a_{n+1} + k\, b_{n+1}).
\end{aligned}
$$

Therefore,

$$\pi_{00}^{n} = \exp(a_n).$$

So, to prove that $\sum_{n=0}^{\infty} \exp(a_n) = \infty$, we will show that $\exp(a_n) \not\to 0$. Consider the function $g(x) = \alpha[\exp(x) - 1]$; note that

$$b_{n+1} = g(b_n) \quad \text{and} \quad b_n = g^{[n-1]}(b_1) = g^{[n-1]}(-\alpha),$$

where $g^{[n]}(\cdot)$ represents the composition of the function g with itself n times and $g^{[1]}(x) = g(x)$. Using the recursive formula for a_{n+1} we obtain

$$a_{n+1} - a_n = \frac{\lambda}{\alpha} b_{n+1} = \frac{\lambda}{\alpha} g^{[n]}(-\alpha).$$

Therefore,

$$\sum_{i=1}^{n-1}(a_{i+1} - a_i) = \frac{\lambda}{\alpha}\sum_{i=1}^{n-1} g^{[i]}(-\alpha) \;\Rightarrow\; a_n - a_1 = \frac{\lambda}{\alpha}\sum_{i=1}^{n-1} g^{[i]}(-\alpha),$$

$$\Rightarrow\; a_n = -\lambda + \frac{\lambda}{\alpha}\sum_{i=1}^{n-1} g^{[i]}(-\alpha).$$

Since $g'(x) = \alpha \exp(x) > 0$ we have that $g(x)$ is an increasing function, thus

$$b_1 = -\alpha < 0 \;\Rightarrow\; b_2 = g(b_1) < g(0) = 0 \;\Rightarrow\; b_{n+1} = g(b_n) < 0 \;\Rightarrow\; b_n < 0 \;\forall n,$$

and it implies that

$$a_{n+1} - a_n = \frac{\lambda}{\alpha} b_{n+1} < 0 \;\Rightarrow\; a_{n+1} < a_n.$$

Since $a_1 = -\alpha < 0$, we can conclude that $(a_n)_{n\geqslant 1}$ is a negative decreasing sequence. On the other hand, from the known inequality $\exp(x) > x + 1 \;\forall x \neq 0$ we have

$$\exp(x) - 1 > x \;\Rightarrow\; \alpha[\exp(x) - 1] > \alpha x \;\Rightarrow\; g(x) > \alpha x, \quad \forall x \neq 0.$$

It is easy to prove by mathematical induction that

$$g^{[n]}(x) > \alpha^n x \quad \forall n \geqslant 1 \text{ and } \forall x \neq 0.$$

Therefore, if $x = -\alpha \neq 0$ then $g^{[n]}(-\alpha) > -\alpha^{n+1}$ and

$$\begin{aligned}
\lim_{n\to\infty} a_n &= \lim_{n\to\infty} \left[-\lambda + \frac{\lambda}{\alpha} \sum_{i=1}^{n-1} g^{[i]}(-\alpha) \right] \\
&= -\lambda + \frac{\lambda}{\alpha} \sum_{i=0}^{\infty} g^{[i]}(-\alpha) \\
&\geqslant -\lambda - \frac{\lambda}{\alpha} \sum_{i=0}^{\infty} \alpha^{i+1} \\
&= -\lambda - \frac{\lambda\alpha^2}{\alpha(1-\alpha)} \\
&= -\frac{\lambda}{1-\alpha}.
\end{aligned}$$

Thus, $\lim_{n\to\infty} \exp(a_n) \geqslant \exp[-\lambda/(1-\alpha)] \neq 0$, i.e., $\sum_{n=1}^{\infty} \exp(a_n)$ diverges. Notice that we proved that state 0 is a recurrent state; now to complete the proof we show that state 0 is a positive recurrent state. Using [Ross [1983], Theorem 4.3.1 (iii) on p. 108] we have

$$\begin{aligned}
\pi_0 &= \lim_{n\to\infty} \pi_{i0}^n \\
&= \lim_{n\to\infty} \pi_{00}^n \\
&= \lim_{n\to\infty} \exp(a_n) > 0.
\end{aligned}$$

Hence 0 is a positive recurrent state and so $\{\pi_j, j = 0, 1, \dots\}$ defined by $\pi_j = \lim_{n\to\infty} \pi_{ij}^n$, exists and defines the unique stationary distribution. □

The next proposition presents a recursive way to find the ordinary moments of the marginal distribution of the INARCH(1) model.

Proposition 4. *Let $(X_t)_{t\geqslant 1}$, $X_t \in \mathbb{N}_0$, be an INARCH(1) process according to equation* (4.2). *Then the $r-$th ordinary moment of the marginal distribution of $(X_t)_{t\geqslant 1}$, is given by*

$$\mathrm{E}[\,X_t^r\,] = \sum_{j=0}^{r}\sum_{i=0}^{j} S(r,j)\binom{j}{i}\alpha^i\lambda^{j-i}\mathrm{E}[\,X_t^i\,],$$

where $S(k,j)$ represents the Stirling number of the second class defined before.

Proof. First, note that if Z is a random variable such that $Z \sim$ Poisson(v), then using Dobiński's formula the $r-$th ordinary moment of Z can be expressed as

$$\mathrm{E}[Z^r] = \sum_{k=0}^{\infty} k^r\,\frac{\exp(-v)v^k}{k!} = \sum_{k=0}^{r} S(r,k)v^k, \tag{4.4}$$

where $S(r,k) = 1/k!\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^r$ is the Stirling number of the second kind, i.e., the number of ways to partition a set of r objects into k non-empty subsets. A quick proof of Dobiński's formula can be found in Pitman [1997].
Then, since $X_t \mid X_{t-1} \sim \mathrm{Po}\big(\alpha X_{t-1}+\lambda\big)$ we can express the $r-$th ordinary moment of X_t as

$$\begin{aligned}
\mathrm{E}[\,X_t^r\,] &= \mathrm{E}\Big\{\mathrm{E}[\,X_t^r \mid X_{t-1}]\Big\}\\
&= \mathrm{E}\left[\sum_{j=0}^{r} S(r,j)\big(\alpha X_{t-1}+\lambda\big)^j\right]\\
&= \sum_{j=0}^{r} S(r,j)\mathrm{E}\Big[(\alpha X_{t-1}+\lambda)^j\Big]\\
&= \sum_{j=0}^{r}\sum_{i=0}^{j} S(r,j)\binom{j}{i}\alpha^j\lambda^{j-i}\mathrm{E}[X_t^i].
\end{aligned}$$

□

The first and second ordinary moments of the INARCH(1) process can be obtained from proposition 4 taking $r = 1$ and $r = 2$, respectively, and using that $S(0,0) = S(1,1) = S(2,1) = S(2,2) = 1$, and $S(0,j) = S(j,0) = 0\ \forall j \neq 0$, then

$$\mathrm{E}[X_t] = \lambda + \alpha\,\mathrm{E}[X_t] \quad \Rightarrow \quad \mathrm{E}[X_t] = \frac{\lambda}{1-\alpha},$$

$$\mathrm{E}[X_t^2] = \lambda + \alpha \mathrm{E}[X_t] + \lambda^2 + 2\alpha\lambda \mathrm{E}[X_t] + \alpha^2 \mathrm{E}[X_t^2] \quad \Rightarrow \quad \mathrm{E}[X_t^2] = \frac{\lambda + \lambda^2 + \alpha\lambda}{1-\alpha}.$$

In Section 2.6 we present three estimation methods in the INARCH(1) model. Albeit the CML estimators have no closed form, we can express the λ estimator in terms of the α estimator, so, it is only necessary to estimate numerically α.

Proposition 5. *If $\widehat{\alpha}_{CML}, \widehat{\lambda}_{CML}$ represent the CML estimators of the α, λ parameters in INARCH(1) model, then*

$$\widehat{\lambda}_{\mathrm{CML}} = \frac{\sum_{t=2}^{T}(x_t - \widehat{\alpha}_{\mathrm{CML}}\, x_{t-1})}{T}.$$

Proof. By solving the system

$$\frac{\partial \ell(\alpha, \lambda; x_1,\, x_2,\, \ldots,\, x_T \,|\, x_1)}{\partial \alpha} = 0,$$
$$\frac{\partial \ell(\alpha, \lambda; x_1,\, x_2,\, \ldots,\, x_T \,|\, x_1)}{\partial \lambda} = 0$$

from equations (2.14) we obtain

$$\sum_{t=2}^{T}\left(\frac{x_t}{\alpha x_{t-1} + \lambda}\right) = T - 1, \tag{4.5}$$

$$\sum_{t=2}^{T}\left(\frac{x_t x_{t-1}}{\alpha x_{t-1} + \lambda}\right) = \sum_{t=2}^{T} x_{t-1}. \tag{4.6}$$

Notice that the general term in equation (4.6) can be written as

$$\begin{aligned} \frac{x_t x_{t-1}}{\alpha x_{t-1} + \lambda} &= \frac{x_t}{\alpha}\left[\frac{(\alpha x_{t-1} + \lambda) - \lambda}{\alpha x_{t-1} + \lambda}\right] \\ &= \frac{x_t}{\alpha}\left(1 - \frac{\lambda}{\alpha x_{t-1} + \lambda}\right) \\ &= \frac{x_t}{\alpha} - \frac{\lambda}{\alpha}\left(\frac{x_t}{\alpha x_{t-1} + \lambda}\right). \end{aligned} \tag{4.7}$$

Therefore, by equations (4.7) and (4.5), we have

$$\begin{aligned} \sum_{t=2}^{T} \frac{x_t x_{t-1}}{\alpha x_{t-1} + \lambda} &= \sum_{t=2}^{T}\left[\frac{x_t}{\alpha} - \frac{\lambda}{\alpha}\left(\frac{x_t}{\alpha x_{t-1} + \lambda}\right)\right] \\ &= \frac{1}{\alpha}\sum_{t=2}^{T} x_t + \frac{\lambda}{\alpha}(T-1). \end{aligned} \tag{4.8}$$

Then using equation (4.8) and (4.6) we obtain

$$\sum_{t=2}^{T} x_t + \lambda(T-1) = \alpha \sum_{t=2}^{T} x_{t-1},$$

thus

$$\widehat{\lambda}_{\text{CML}} = \frac{\sum_{t=2}^{T}(\widehat{\alpha}_{\text{CML}} x_{t-1} - x_t)}{T-1}.$$

□

This result was also proved by Weiß [2010].

4.3 Forecasting for the INARCH(1) process

To the best of our knowledge, forecasting for INARCH(1) processes has not been studied. In this section we propose one two and $h-$steps ahead forecasts for the INARCH(1) process and we find and prove its properties.

The one-step ahead conditional distribution of the INARCH(1) process is a Poisson distribution. Thus, considering the minimum mean squared error and the minimum mean absolute error as optimization criteria to obtain forecasts one-step ahead, we propose the conditional mean and the conditional median as forecasts one-step ahead. The mean of a Poisson distribution can be a non-integer value; then, we have to use its integer part as forecast. Further, if Z is a random variable with Poisson distribution with parameter v then

$$\text{mode } (Z) = \begin{cases} \lfloor v \rfloor, & \text{if} \quad v \in \mathbb{N}, \\ v-1 \text{ and } v, & \text{if} \quad v \notin \mathbb{N}. \end{cases}$$

In fact, note that

$$\frac{\mathrm{P}(Z=k+1)}{P(Z=k)} = \frac{\exp(-v)v^{k+1}/(k+1)!}{\exp(-v)v^k/k!} = \frac{v}{k+1},$$

thus,

$$\mathrm{P}(Z=k) \geqslant \mathrm{P}(Z=k+1) \quad \text{for every} \quad k+1 \geqslant v, \quad \text{and,}$$

$$P(Z = k) \geqslant P(Z = k-1) \quad \text{for every} \quad k \leqslant v,$$

then $P(Z = k) \geqslant P(Z = k+1)$ and $P(Z = k) \geqslant P(Z = k-1)$ for every k such as $v - 1 \leqslant k \leqslant v$. Therefore, if k_v denotes the mode of Z we have

i. If $v < 1$ then $k_v = 0$

ii. If $v > 1$ and $v \notin \mathbb{N}$ then $k_v = \lfloor v \rfloor$

iii. If $v \in \mathbb{N}$ then $k_v = v - 1$ and $k_v = v$ are both modes.

Hence, when the parameter of a Poisson distribution is non-integer, the integer part of the mean is equal to the mode.

Our purpose is to provide a coherent forecast of a future value x_{T+h}, i. e., integer-valued forecast, given that we have observed the series up to time T, i.e., $x_1, x_2, \ldots, x_T$ are known. It is natural to use the $h-$steps ahead conditional distribution to provide forecasts. We know the conditional distribution of $X_{T+1} \mid X_T$, but we do not know the conditional distribution of $X_{T+h} \mid X_T$ for $h \geqslant 2$. However, we obtained a closed expression for the conditional probability generating function (PGF) of $X_{T+h} \mid X_T$, which allows to find the probabilities. Also, we got expressions for the conditional mean and variance of X_{T+h} given X_T, respectively.

Proposition 6. *In the INARCH(1) model the $h-$steps conditional mean is given by*

$$\mathrm{E}\left[\, X_{T+h} \mid X_T \,\right] = \alpha^h X_T + \left(\frac{1-\alpha^h}{1-\alpha} \right) \lambda.$$

Proof. We prove the proposition by mathematical induction.

- ⋆ For $k = 1$ we have

$$\mathrm{E}[X_{T+1} \mid X_T] = \mathrm{E}[\, (\alpha * X_T + \epsilon_{T+1}) | X_T \,] = \alpha X_T + \lambda$$

- ⋆ Suppose for $k = h$ that $\mathrm{E}\left[\, X_{T+h} \mid X_T \,\right] = \alpha^h X_T + \left(\frac{1-\alpha^h}{1-\alpha} \right) \lambda$

⋆ For $k = h+1$ we have

$$\begin{aligned}
\mathrm{E}[\,X_{T+h+1}\,|\,X_T] &= \mathrm{E}[\alpha * X_{T+h} + \epsilon_{T+h+1}|X_T] \\
&= \mathrm{E}[\alpha * X_{T+h}|X_T] + \mathrm{E}[\epsilon_{T+h+1}|X_T] \\
&= \mathrm{E}\{\mathrm{E}[\alpha * X_{T+h}|X_{T+h}]|X_T\} + \mathrm{E}[\epsilon_{T+h+1}] \\
&= \mathrm{E}[\,\alpha X_{T+h}\,|\,X_T\,] + \lambda \\
&= \alpha\left[\alpha^h X_T + \left(\frac{1-\alpha^h}{1-\alpha}\right)\lambda\right] + \lambda \\
&= \alpha^{h+1} X_T + \left(\frac{\alpha-\alpha^{h+1}}{1-\alpha} + 1\right)\lambda \\
&= \alpha^{h+1} X_T + \left(\frac{1-\alpha^{h+1}}{1-\alpha}\right)\lambda
\end{aligned}$$

□

Although proposition above and Equation 2.5 provide the same expression for the $h-$step conditional mean for the Poisson INAR(1) and the INARCH(1) processes, the proof presented by Freeland [1998] use an argument which is not true for the INARCH(1) process.

Proposition 7. *In the INARCH(1) model the $h-$steps conditional variance is given by*

$$\mathrm{Var}[\,X_{T+h}\,|\,X_T\,] = \alpha^h X_T\left(\frac{1-\alpha^h}{1-\alpha}\right) + \lambda\left(\frac{1-\alpha^{2h-1}}{1-\alpha}\right) + \lambda\,\mathrm{P}_h(\alpha),$$

where

$$\mathrm{P}_h(\alpha) = \alpha^2\left(\frac{1-\alpha^{h-2}}{1-\alpha} + \mathrm{P}_{h-1}(\alpha)\right), \quad \textit{for } h \geqslant 3,$$

and $\mathrm{P}_1(\alpha) = \mathrm{P}_2(\alpha) = 0$.

Proof. We prove the proposition by mathematical induction.

⋆ For $k = 1$ we have

$$\mathrm{Var}[\,X_{T+1}\,|\,X_T\,] = \mathrm{Var}[\,\alpha * X_T + \epsilon_{T+1}\,|\,X_T\,]$$

$$= \mathrm{Var}\left[\, \alpha * X_T \mid X_T \,\right] + \mathrm{Var}\left[\, \epsilon_{T+1} \mid X_T \,\right]$$
$$= \alpha X_T + \lambda,\ \mathrm{P}_1(\alpha) = 0.$$

* ⋆ For $k = 2$ we have

$$\begin{aligned}
\mathrm{Var}\left[\, X_{T+2} \mid X_T \,\right] &= \mathrm{Var}\left[\, \alpha * X_{T+1} + \epsilon_{T+2} \mid X_T \,\right] \\
&= \mathrm{Var}\big[\mathrm{E}\big(\alpha * X_{T+1}\big|X_{T+1}\big)\big|X_T\big] + \mathrm{E}\big[\mathrm{Var}\big(\alpha * X_{T+1}\big|X_{T+1}\big)\big|X_T\big] + \lambda \\
&= \mathrm{Var}\big[\alpha X_{T+1}\big|X_T\big] + \mathrm{E}\big[\alpha X_{T+1}\big|X_T\big] + \lambda \\
&= \alpha^2(\alpha X_T + \lambda) + \alpha(\alpha X_T + \lambda) + \lambda \\
&= \alpha^2 X_T \left(\frac{1-\alpha^2}{1-\alpha}\right) + \lambda\left(\frac{1-\alpha^3}{1-\alpha}\right),\ \mathrm{P}_2(\alpha) = 0.
\end{aligned}$$

* ⋆ Suppose for $k = h$ that

$$\mathrm{Var}\left[\, X_{T+h} \mid X_T \,\right] = \alpha^h X_T \left(\frac{1-\alpha^h}{1-\alpha}\right) + \lambda\left(\frac{1-\alpha^{2h-1}}{1-\alpha}\right) + \lambda\, \mathrm{P}_h(\alpha).$$

* ⋆ For $k = h+1$ we have

$$\begin{aligned}
\mathrm{Var}\left[\, X_{T+h+1} \mid X_T \,\right] &= \mathrm{Var}\left[\, \alpha * X_{T+h} + \epsilon_{T+h+1} \mid X_T \,\right] \\
&= \mathrm{Var}\big[\mathrm{E}\big(\alpha * X_{T+h}\big|X_{T+h}\big)\big|X_T\big] + \mathrm{E}\big[\mathrm{Var}\big(\alpha * X_{T+h}\big|X_{T+h}\big)\big|X_T\big] + \lambda \\
&= \alpha^2\left[\alpha^h X_T \left(\frac{1-\alpha^h}{1-\alpha}\right) + \lambda\left(\frac{1-\alpha^{2h-1}}{1-\alpha}\right) + \lambda\, \mathrm{P}_h(\alpha)\right] \\
&\quad + \alpha^{h+1} X_T + \lambda\left(\frac{1-\alpha^{h+1}}{1-\alpha}\right) \\
&= \alpha^{h+1} X_T \left(\frac{1-\alpha^{h-1}}{1-\alpha}\right) + \lambda\left(\frac{1-\alpha^{2h+1}}{1-\alpha}\right) \\
&\quad + \alpha^2\lambda\left(\frac{1-\alpha^{h-1}}{1-\alpha}\right) + \alpha^2\lambda\, \mathrm{P}_h(\alpha) \\
&= \alpha^{h+1} X_T \left(\frac{1-\alpha^{h-1}}{1-\alpha}\right) + \lambda\left(\frac{1-\alpha^{2h+1}}{1-\alpha}\right) + \lambda\, \mathrm{P}_{h+1}(\alpha).
\end{aligned}$$

□

Proposition 8. *As h goes to infinity, the conditional mean and variance of the INARCH(1) model converge to the unconditional mean and variance, respectively. We have*

$$\lim_{h\to\infty} \mathrm{E}\left[\, X_{T+h} \mid X_T \,\right] = \frac{\lambda}{1-\alpha},$$

$$\lim_{h\to\infty} \text{Var}\left[\, X_{T+h} \,|\, X_T \,\right] = \frac{\lambda}{(1-\alpha)(1-\alpha^2)}.$$

Proof. It is easy to see that $\lim_{h\to\infty} \text{E}\left[\, X_{T+h} \,|\, X_T \,\right] = \frac{\lambda}{1-\alpha}$ and $\lim_{h\to\infty} \text{P}_h(\alpha) = \frac{\alpha^2}{(1-\alpha)(1-\alpha^2)}$, hence,

$$\begin{aligned}
\lim_{h\to\infty} \text{Var}[X_{T+h}|X_T] &= \frac{\lambda}{1-\alpha} + \lambda \lim_{h\to\infty} \text{P}_h(\alpha) \\
&= \frac{\lambda}{1-\alpha} + \lambda \frac{\alpha^2}{(1-\alpha)(1-\alpha^2)} \\
&= \frac{\lambda}{(1-\alpha)(1-\alpha^2)}.
\end{aligned}$$

□

Proposition 9. *For the INARCH(1) model, the $h-$steps ahead conditional moment generating function (MGF) is given by*

$$\mathcal{M}_{X_{T+h}|X_T}(s) = \exp\left\{ \frac{1}{\alpha} \left[\lambda \sum_{i=1}^{h-1} g^{[i]}(s) + (\alpha X_T + \lambda) g^{[h]}(s) \right] \right\}, \tag{4.9}$$

where $g(s) = \alpha[\exp(s) - 1]$, $g^{[i]}(\cdot)$ represents the composition of the function g with itself i times and $g^{[1]}(s) = g(s)$.

Proof. We prove the proposition by mathematical induction.

- ⋆ For $k = 1$

$$\begin{aligned}
\mathcal{M}_{X_{T+1}|X_T}(s) &= E\left[\exp(\alpha * X_T + \epsilon_{T+1})|X_T\right] \\
&= \exp\left[(\alpha X_T + \lambda)(\exp(s) - 1)\right] \\
&= \exp\left[\frac{1}{\alpha}(\alpha X_T + \lambda) g(s)\right]
\end{aligned}$$

- ⋆ For $k = h$, suppose

$$\mathcal{M}_{X_{T+h}|X_T}(s) = \exp\left\{ \frac{1}{\alpha} \left[\lambda \sum_{i=1}^{h-1} g^{[i]}(s) + (\alpha X_T + \lambda) g^{[h]}(s) \right] \right\};$$

- ⋆ then, for $k = h + 1$ we have

$$\begin{aligned}
\mathcal{M}_{X_{T+h+1}|X_T}(s) &= E\big[\exp(\alpha * X_{T+h} + \epsilon_{T+h+1})|X_T\big] \\
&= \mathrm{E}\{\mathrm{E}\big[\exp(\alpha * X_{T+h} + \epsilon_{T+h+1})\big|X_{T+h}\big]\big|X_T\} \\
&= \mathrm{E}\{\exp[(\alpha X_{T+h} + \lambda)(\exp(s) - 1)]\big|X_T\} \\
&= \exp\left[\frac{\lambda}{\alpha} g(s)\right] \mathcal{M}_{X_{T+h}|X_T}(g(s)) \\
&= \exp\left[\frac{\lambda}{\alpha} g(s)\right] \exp\left\{\frac{1}{\alpha}\left[\lambda \sum_{i=1}^{h-1} g^{[i]}(g(s)) + (\alpha X_T + \lambda) g^{[h]}(g(s))\right]\right\} \\
&= \exp\left\{\frac{1}{\alpha}\left[\lambda g(s) + \lambda \sum_{i=1}^{h-1} g^{[i+1]}(s) + (\alpha X_T + \lambda) g^{[k+1]}(s)\right]\right\} \\
&= \exp\left\{\frac{1}{\alpha}\left[\lambda \sum_{i=1}^{h} g^{[i]}(s) + (\alpha X_T + \lambda) g^{[k+1]}(s)\right]\right\}.
\end{aligned}$$

□

Note that we can obtain the conditional PGF from the conditional MGF; we have $\mathcal{G}_{X_{T+1}|X_T}(s) = \mathcal{M}_{X_{T+1}|X_T}(\log(s))$, therefore using equation (4.9), we obtain

$$\mathcal{G}_{X_{T+h}|X_T}(s) = \exp\left\{\frac{1}{\alpha}\left[\lambda \sum_{i=1}^{h-1} g^{[i]}(\log(s)) + (\alpha X_T + \lambda) g^{[h]}(\log(s))\right]\right\}. \tag{4.10}$$

4.4 Forecasting one-step ahead

Now, we focus on the forecast one-step ahead. Although we know the conditional distribution of $X_{T+1} \mid X_T$, there is no expression for the median of a Poisson distribution. However, there are some results on its bounds.

Let Z be a random variable, $Z \sim \text{Po}\,(v)$. Chen and Rubin [1986] conjectured the inequality

$$v - \log(2) \leqslant \text{median}(Z) < v + \frac{1}{3}. \tag{4.11}$$

After that, Choi [1994] proved it and Adell and Jodrá [2005] proved that the bounds in (4.11) are the best possible absolute bounds. Notice that in the interval

$$\left[v - \log(2)\, ,\, v + \frac{1}{3} \right)$$

there can be no more than two integers, namely, $\lceil v - \log(2) \rceil$ and $\left\lfloor v + \frac{1}{3} \right\rfloor$, thus, if $\lceil v - \log(2) \rceil \neq \left\lfloor v + \frac{1}{3} \right\rfloor$ we have that

$$\text{median}(Z) = \lceil v - \log(2) \rceil \quad \text{or} \quad \text{median}(Z) = \left\lfloor v + \frac{1}{3} \right\rfloor .$$

Assuming known parameters and using equation (4.11) we obtain the inequality

$$\alpha X_T + \lambda - \log(2) \leqslant X_T^{(\mathrm{Mn})}(1) < \alpha X_T + \lambda + \frac{1}{3},$$

where $X_T^{(\mathrm{Mn})}(1)$ represents the median of the Poisson distribution with mean $\alpha X_T + \lambda$. Initially, based on the above result, we have approximated $X_T^{(\mathrm{Mn})}(1)$ as below

$$X_T^{(\mathrm{Mn})}(1) \approx \begin{cases} 0, & \text{if} \quad \alpha X_T + \lambda \leqslant \log(2), \\ 1, & \text{if} \quad \log(2) < \alpha X_T + \lambda \leqslant \frac{5}{3}, \\ m, & \text{if} \quad m - \frac{1}{3} < \alpha X_T + \lambda \leqslant m + \frac{2}{3}, \text{ for } m \geqslant 2 \end{cases}$$

$$= \begin{cases} 0, & \text{if} \quad \alpha X_T + \lambda \leqslant \log(2), \\ 1, & \text{if} \quad \log(2) < \alpha X_T + \lambda \leqslant \frac{5}{3}, \\ \left\lceil \alpha X_T + \lambda - \frac{2}{3} \right\rceil, & \text{if} \quad \alpha X_T + \lambda > \frac{5}{3}; \end{cases}$$

then, we observed that we can approximate the median by a more compact expression obtaining almost the same results. This can be confirmed by the simulation study which we are presenting in the next session. So, we propose to approximate the exact median $X_T^{(\mathrm{Mn})}(1)$ of $X_{T+1} \mid X_T$ by $X_T^{(\mathrm{Map})}(1)$

$$X_T^{(\mathrm{Mn})}(1) \approx X_T^{(\mathrm{Map})}(1) = \left\lceil \alpha X_T + \lambda - \frac{2}{3} \right\rceil ,$$

and, for unknown parameters, we propose to approximate the exact median $X_T^{(\mathrm{Mn})}(1)$ by $X_T^{(\mathrm{Map})}(1)$

$$X_T^{(\mathrm{Map})}(1) = \left\lceil \widehat{\alpha} X_T + \widehat{\lambda} - \frac{2}{3} \right\rceil , \tag{4.12}$$

where $\widehat{\alpha}$ and $\widehat{\lambda}$ represent estimators of α and λ respectively.

For the follow propositions: Proposition (10), Proposition (11), Proposition (12) and Proposition (14) we suppose that for large T if $\widehat{\alpha}$ and $\widehat{\lambda}$ represent estimators of α and λ respectively, the conditional distribution of $\left(\widehat{\alpha}, \widehat{\lambda} \,\middle|\, X_T\right)$ is also the distribution of $\left(\widehat{\alpha}, \widehat{\lambda}\right)$.

Proposition 10. *Let $\widehat{\alpha}$ and $\widehat{\lambda}$ be estimators of α and λ such that*

$$\sqrt{T}\begin{pmatrix}\widehat{\alpha}-\alpha\\ \widehat{\lambda}-\lambda\end{pmatrix} \xrightarrow{D} \mathcal{N}_2(\mathbf{0},\Sigma).$$

Then

$$\lim_{T\to\infty} \mathrm{E}\left[X_T^{(\mathrm{Map})}(1)\middle|X_T\right] = \left\lceil \alpha X_T + \lambda - \frac{2}{3}\right\rceil.$$

Proof. We have that

$$\begin{aligned}
\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)\middle|X_T\right] &= \sum_{k=0}^{\infty} \mathrm{P}\left(X_T^{(\mathrm{Map})}(1) > k\middle|X_T\right)\\
&= \sum_{k=0}^{\infty} \mathrm{P}\left(\left\lceil \widehat{\alpha}X_T + \widehat{\lambda} - \frac{2}{3}\right\rceil > k \,\middle|\, X_T\right)\\
&= \sum_{k=0}^{\infty} \mathrm{P}\left(\widehat{\alpha}X_T + \widehat{\lambda} - \frac{2}{3} > k \,\middle|\, X_T\right)\\
&= \sum_{k=1}^{\infty} \mathrm{P}\left(\widehat{\alpha}X_T + \widehat{\lambda} > k - \frac{1}{3} \,\middle|\, X_T\right).
\end{aligned}$$

For large T we have approximately

$$\widehat{\alpha}X_T + \widehat{\lambda} \,|\, X_T \sim \mathcal{N}\left(\nu, \frac{\sigma^2_{\alpha,\lambda}}{T}\right),$$

where $\nu = \alpha X_T + \lambda$ and $\sigma^2_{\alpha,\lambda}$ depends on α and λ. Thus,

$$\begin{aligned}
\mathrm{E}\left[X_T^{(\mathrm{Map})}(1) \,\middle|\, X_T\right] &= \sum_{k=1}^{\infty} \mathrm{P}\left(\widehat{\alpha}X_T + \widehat{\lambda} > k - \frac{1}{3} \,\middle|\, X_T\right)\\
&\approx \sum_{k=1}^{\infty} \mathrm{P}\left(\mathcal{Z} > \frac{\sqrt{T}(k - \frac{1}{3} - \nu)}{\sigma_{\alpha,\lambda}} \,\middle|\, X_T\right)
\end{aligned}$$

$$= \sum_{k=1}^{\infty} \left[1 - \Phi\left(\frac{\sqrt{T}(k - \frac{1}{3} - \nu)}{\sigma_{\alpha,\lambda}} \right) \right],$$

where $\mathcal{Z}$ is a random variable with standard normal distribution and $\Phi(\cdot)$ denotes the cumulative distribution function of $\mathcal{Z}$. The approximation is true from the Central Limit Theorem. Considering $h(x) = \log\left[\frac{\Phi(x)}{1-\Phi(x)}\right]$ and using the Taylor series expansion of $h(x)$ at 0, Tocher [1963] found an approximation for $\mathrm{P}(\mathcal{Z} > z)$; the author obtained

$$1 - \Phi(z) \approx \frac{1}{1 + \exp(cz)},$$

where $c = 2\sqrt{2/\pi}$, therefore

$$\begin{aligned}
\mathrm{E}\left[X_T^{(\mathrm{Map})}(1) \,\middle|\, X_T \right] &\approx \sum_{k=1}^{\infty} \left[1 - \Phi\left(\frac{\sqrt{T}(k - \frac{1}{3} - \nu)}{\sigma_{\alpha,\lambda}} \right) \right] \\
&\approx \sum_{k=1}^{\infty} \frac{1}{1 + \exp\left[c\left(\sqrt{T}(k - \frac{1}{3} - \nu)/\sigma_{\alpha,\lambda} \right) \right]} \\
&= \frac{1}{\exp[-c\sqrt{T}(\nu + 1/3)/\sigma_{\alpha,\lambda}]} \times \\
&\quad \sum_{k=1}^{\infty} \frac{1}{\exp[c\sqrt{T}(\nu + 1/3)/\sigma_{\alpha,\lambda}] + [\exp(c\sqrt{T}/\sigma_{\alpha,\lambda})]^k}.
\end{aligned}$$

If $b = \exp(c\sqrt{T}/\sigma_{\alpha,\lambda})$ and $a = b^{\nu+1/3}$, we can write the last expression as

$$\mathrm{E}\left[X_T^{(\mathrm{Map})}(1) \,\middle|\, X_T \right] \approx \frac{1}{a^{-1}} \sum_{k=1}^{\infty} \frac{1}{a + b^k}. \tag{4.13}$$

Notice that

$$\sum_{k=1}^{\infty} \frac{1}{a + b^k} = \sum_{k=1}^{N-1} \frac{1}{a + b^k} + \sum_{k=N}^{\infty} \frac{1}{a + b^k}, \tag{4.14}$$

where we choose $N = \lceil \nu + 1/3 \rceil$, so, $k \geqslant \nu + 1/3$ for each k in the second sum of right side. Using Taylor expansion again to approximate $h_k(x) = \dfrac{1}{x + b^k}$, we obtain

$$h_k(a) = \sum_{i=0}^{m} \frac{(-a)^i}{b^{k(i+1)}} + \frac{(-a)^{m+1}}{(\xi + b^k)^{m+2}},$$

where $0 < \xi < a$. Then,

$$\sum_{k=1}^{\infty} \frac{1}{a + b^k} = \sum_{k=1}^{N-1} \frac{1}{a + b^k} + \sum_{k=N}^{\infty} \left[\sum_{i=0}^{m} \frac{(-a)^i}{b^{k(i+1)}} + \frac{(-a)^{m+1}}{(\xi + b^k)^{m+2}} \right]$$

$$\begin{aligned}
&\approx \sum_{k=1}^{N-1} \frac{1}{a+b^k} + \sum_{i=0}^{m} (-a)^i \sum_{k=N}^{\infty} \left(\frac{1}{b^{i+1}} \right)^k \\
&= \sum_{k=1}^{N-1} \frac{1}{a+b^k} + \sum_{i=0}^{m} \frac{(-1)^i \, b^{\, i(\nu+1/3)} \, b^{\, i+1}}{b^{\, N(i+1)} \, (b^{\, i+1} - 1)} \\
&= \sum_{k=1}^{N-1} \frac{1}{a+b^k} + \sum_{i=0}^{m} \frac{(-1)^i \, b^{\, i(\nu+1/3)}}{b^{\, N(i+1)} - b^{(N-1)(i+1)}},
\end{aligned}$$

where

$$\lim_{T\to\infty} \frac{(-a)^{m+1}}{(\xi+b^k)^{m+2}} = 0. \tag{4.15}$$

In fact, if $d = \exp(c/\sigma_{\alpha,\lambda})$ then $d > 1, b = d^{\sqrt{T}}$ and $a = d^{\sqrt{T}(\nu+1/3)}$. So, we have

$$\begin{aligned}
\lim_{T\to\infty} \frac{(-a)^{m+1}}{(\xi+b^k)^{m+2}} &= \lim_{T\to\infty} \frac{(-1)^{m+1} d^{\sqrt{T}(\nu+1/3)(m+1)}}{(\xi + d^{\sqrt{T}k})^{m+2}} \\
&= (-1)^{m+1} \lim_{T\to\infty} \left[\left(\frac{d^{\sqrt{T}(\nu+1/3)}}{\xi + d^{\sqrt{T}k}} \right)^{m+1} \left(\frac{1}{\xi + d^{\sqrt{T}\,k}} \right) \right] \\
&= (-1)^{m+1} \Bigg(\underbrace{\lim_{T\to\infty} \frac{d^{\sqrt{T}(\nu+1/3-k)}}{\xi \, d^{-\sqrt{T}k} + 1}}_{=0} \Bigg)^{m+1} \underbrace{\lim_{T\to\infty} \frac{1}{\xi + d^{\sqrt{T}\,k}}}_{=0} \\
&= 0.
\end{aligned}$$

The first limit is equal zero because $k \geqslant N = \lceil \nu + 1/3 \rceil \geqslant \nu + 1/3$, thus $\nu + 1/3 - k \leqslant 0$.

For $m = 1$ we can approximate the right side of equation (4.14) as

$$\sum_{k=1}^{\infty} \frac{1}{a+b^k} \approx \sum_{k=1}^{N-1} \frac{1}{a+b^k} + \frac{1}{b^N - b^{N-1}} - \frac{b^{\nu+1/3}}{b^{2N} - b^{2(N-1)}},$$

then, equation (4.13) can be expressed as

$$\begin{aligned}
\mathrm{E}\left[X_T^{(\mathrm{Map})}(1) \,\middle|\, X_T \right] &\approx \frac{1}{b^{-(\nu+1/3)}} \left[\sum_{k=1}^{N-1} \frac{1}{a+b^k} + \frac{1}{b^N - b^{N-1}} - \frac{b^{\nu+1/3}}{b^{2N} - b^{2(N-1)}} \right] \\
&= \sum_{k=1}^{N-1} \frac{1}{1 + b^{k-\nu-1/3}} + \frac{1}{b^{N-\nu-1/3} - b^{N-\nu-4/3}} - \frac{1}{b^{2(N-\nu-1/3)} - b^{2(N-\nu-4/3)}}.
\end{aligned} \tag{4.16}$$

Replacing $b = d^{\sqrt{T}}$ Equation (4.16) becomes

$$\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)\,\middle|\,X_T\right] \approx \underbrace{\sum_{k=1}^{N-1}\frac{1}{1+d^{\sqrt{T}(k-\nu-1/3)}}}_{k<N=\lceil\nu+1/3\rceil \Rightarrow k-\nu-1/3<0} + \underbrace{\frac{d^{\sqrt{T}(\nu+1/3-N)}}{1-d^{-\sqrt{T}}} - \frac{d^{2\sqrt{T}(\nu+1/3-N)}}{1-d^{-2\sqrt{T}}}}_{N=\lceil\nu+1/3\rceil\geqslant\nu+1/3 \Rightarrow \nu+1/3-N\leqslant 0}.$$

Therefore,

$$\lim_{T\to\infty}\left[\sum_{k=1}^{N-1}\frac{1}{1+d^{\sqrt{T}(k-\nu-1/3)}}\right] = N-1$$

and

$$\lim_{T\to\infty}\left[\frac{d^{\sqrt{T}(\nu+1/3-N)}}{1-d^{-\sqrt{T}}} - \frac{d^{2\sqrt{T}(\nu+1/3-N)}}{1-d^{-2\sqrt{T}}}\right] = 0.$$

Then

$$\begin{aligned}\lim_{T\to\infty}\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)\,\middle|\,X_T\right] &= \lceil\nu+1/3\rceil-1\\ &= \lceil\nu-2/3+1\rceil-1\\ &= \lceil\nu-2/3\rceil\\ &= \left\lceil\alpha X_T+\lambda-\frac{2}{3}\right\rceil.\end{aligned}$$

□

Proposition 11. *Let $\widehat{\alpha}$ and $\widehat{\lambda}$ be estimators of α and λ such that*

$$\sqrt{T}\begin{pmatrix}\widehat{\alpha}-\alpha\\ \widehat{\lambda}-\lambda\end{pmatrix} \xrightarrow{D} \mathcal{N}_2(\mathbf{0},\Sigma).$$

Then

$$\lim_{T\to\infty}\mathrm{Var}\left[X_T^{(\mathrm{Map})}(1)\middle|X_T\right] = 0.$$

Proof. By the variance definition and the last proposition it is sufficient to prove that

$$\lim_{T\to\infty}\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)^2\,\middle|\,X_T\right] = \left(\left\lceil\alpha X_T+\lambda-\frac{2}{3}\right\rceil\right)^2.$$

This demonstration is very similar to the demonstration of proposition 10. It is known that $\sum_{k=1}^{j}(2k-1) = j^2$, then

$$\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)^2\,\middle|\,X_T\right] = \sum_{j=1}^{\infty} j^2\,\mathrm{P}\left(X_T^{(\mathrm{Map})}(1)=j\,\middle|\,X_T\right)$$

$$
\begin{aligned}
&= \sum_{j=1}^{\infty} \mathrm{P}\Big(X_T^{(\mathrm{Map})}(1) = j \,\Big|\, X_T\Big) \sum_{k=1}^{j} (2k-1) \\
&= \sum_{k=1}^{\infty} (2k-1) \sum_{j=k}^{\infty} \mathrm{P}\left(X_T^{(\mathrm{Map})}(1) = j \,\Big|\, X_T\right) \\
&= \sum_{k=1}^{\infty} (2k-1)\, \mathrm{P}\left(X_T^{(\mathrm{Map})}(1) \geqslant k \,\Big|\, X_T\right), \\
&= \sum_{k=1}^{\infty} (2k-1) \mathrm{P}\left(\left\lceil \widehat{\alpha} X_T + \widehat{\lambda} - \frac{2}{3} \right\rceil \geqslant k \,\Big|\, X_T\right) \\
&= \sum_{k=1}^{\infty} (2k-1) \mathrm{P}\left(\widehat{\alpha} X_T + \widehat{\lambda} - \frac{2}{3} > k-1 \,\Big|\, X_T\right) \\
&= \sum_{k=1}^{\infty} (2k-1) \mathrm{P}\left(\widehat{\alpha} X_T + \widehat{\lambda} > k - \frac{1}{3} \,\Big|\, X_T\right).
\end{aligned}
$$

Thus, by the normality hypothesis and the normal approximation of Tocher [1963], we have

$$
\begin{aligned}
\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)^2 \,\Big|\, X_T\right] &= \sum_{k=1}^{\infty} (2k-1) \mathrm{P}\left(\widehat{\alpha} X_T + \widehat{\lambda} > k - \frac{1}{3} \,\Big|\, X_T\right) \\
&\approx \sum_{k=1}^{\infty} (2k-1) \mathrm{P}\left(\mathcal{Z} > \frac{\sqrt{T}(k - \frac{1}{3} - \nu)}{\sigma_{\alpha,\lambda}} \,\Big|\, X_T\right) \\
&= \sum_{k=1}^{\infty} (2k-1) \left[1 - \Phi\left(\frac{\sqrt{T}(k - \frac{1}{3} - \nu)}{\sigma_{\alpha,\lambda}}\right)\right] \\
&\approx \sum_{k=1}^{\infty} \frac{(2k-1)}{1 + \exp\left[c\left(\sqrt{T}(k - \frac{1}{3} - \nu)/\sigma_{\alpha,\lambda}\right)\right]},
\end{aligned}
$$

where $\mathcal{Z}$ is a random variable with standard normal distribution, $\Phi(\cdot)$ denotes the cumulative distribution function of $\mathcal{Z}$ and $c = 2\sqrt{2/\pi}$. If $b = \exp(c\sqrt{T}/\sigma_{\alpha,\lambda})$ and $a = b^{\nu+1/3}$, we can write the last expression as

$$
\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)^2 \,\Big|\, X_T\right] \approx \frac{1}{a^{-1}} \sum_{k=1}^{\infty} \frac{2\,k-1}{a+b^k}. \tag{4.17}
$$

Also

$$
\sum_{k=1}^{\infty} \frac{2k-1}{a+b^k} = \sum_{k=1}^{N-1} \frac{2k-1}{a+b^k} + \sum_{k=N}^{\infty} \frac{2k-1}{a+b^k}, \tag{4.18}
$$

where we choose $N = [\nu + 1/3]$. Using Taylor expansion to approximate $h_k(x) = \dfrac{2k-1}{x+b^k}$, we obtain

$$h_k(a) = \sum_{i=0}^{m} \frac{(-a)^i(2k-1)}{b^{k(i+1)}} + \frac{(-a)^{m+1}(2k-1)}{(\xi + b^k)^{m+2}},$$

where $0 < \xi < a$, then,

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{2k-1}{a+b^k} &= \sum_{k=1}^{N-1} \frac{2k-1}{a+b^k} + \sum_{k=N}^{\infty} \left[\sum_{i=0}^{m} \frac{(-a)^i(2k-1)}{b^{k(i+1)}} + \frac{(-a)^{m+1}(2k-1)}{(\xi+b^k)^{m+2}} \right] \\
&\approx \sum_{k=1}^{N-1} \frac{2k-1}{a+b^k} + \sum_{i=0}^{m} (-a)^i \sum_{k=N}^{\infty} \frac{2k-1}{(b^{i+1})^k} \\
&= \sum_{k=1}^{N-1} \frac{1}{a+b^k} + \sum_{i=0}^{m} \frac{(-1)^i(3-2N)b^{(\nu-N+4/3)i-(N-1)}}{(b^{i+1}-1)^2} \\
&\quad + \sum_{i=0}^{m} \frac{(-1)^i(2N-1)b^{(\nu-N+7/3)i-(N-2)}}{(b^{i+1}-1)^2},
\end{aligned}$$

where

$$\lim_{T\to\infty} \frac{(-a)^{m+1}(2k-1)}{(\xi+b^k)^{m+2}} = (-1)^{m+1}(2k-1) \lim_{T\to\infty} \frac{a^{m+1}}{(\xi+b^k)^{m+2}} = 0,$$

from Equation (4.15).

For $m = 1$ we can approximate the right side of equation (4.18) as

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{2k-1}{a+b^k} \approx \sum_{k=1}^{N-1} \frac{2k-1}{a+b^k} &+ \frac{(3-2N)b^{-(N-1)} + (2N-1)b^{-(N-2)}}{(b-1)^2} \\
&- \frac{(3-2N)b^{\nu-2N+7/3} + (2N-1)b^{\nu-2N+13/3}}{(b^2-1)^2}.
\end{aligned}$$

Then, equation (4.17) can be expressed as

$$\begin{aligned}
\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)^2 \,\middle|\, X_T \right] \approx \sum_{k=1}^{N-1} \frac{2k-1}{1+b^{k-(\nu+1/3)}} &+ \frac{(3-2N)b^{(\nu+1/3)-(N-1)}}{(b-1)^2} \\
&+ \frac{(2N-1)b^{(\nu+1/3)-(N-2)}}{(b-1)^2} \\
&- \frac{(3-2N)b^{2(\nu-N+4/3)} + (2N-1)b^{2(\nu-N+7/3)}}{(b^2-1)^2}.
\end{aligned} \tag{4.19}$$

Since we chose $N = \lceil \nu + 1/3 \rceil$, we have

$$N - 1 < \nu + 1/3 \leqslant N \;\Rightarrow\; 0 < \nu + 1/3 - (N-1) \leqslant 1.$$

If we consider $s = \nu + 1/3 - (N-1)$ and $d = \exp(c/\sigma_{\alpha,\lambda})$, then using the last inequality $0 < s \leqslant 1,\; b = d^{\sqrt{T}}$ and equation (4.19) becomes

$$\begin{aligned}
\mathrm{E}\left[\mathrm{X}_T^{(\mathrm{Map})}(1)^2 \,\middle|\, X_T \right] \approx{}& \underbrace{\sum_{k=1}^{N-1} \frac{2k-1}{1 + d^{\sqrt{T}[k-(\nu+1/3)]}}}_{k<N=\lceil \nu+1/3 \rceil \;\Rightarrow\; k-(\nu+1/3)<0} + \frac{(3-2N)d^{s\sqrt{T}} + (2N-1)d^{(s+1)\sqrt{T}}}{(d^{\sqrt{T}}-1)^2} \\
&- \frac{(3-2N)d^{2\,s\sqrt{T}} + (2N-1)d^{2\,(s+1)\sqrt{T}}}{(d^{2\sqrt{T}}-1)^2}.
\end{aligned}$$

It is easy to see that

$$\lim_{T\to\infty} \left[\frac{(3-2N)d^{s\sqrt{T}} + (2N-1)d^{(s+1)\sqrt{T}}}{(b-1)^2} - \frac{(3-2N)d^{2\,s\sqrt{T}} + (2N-1)d^{2\,(s+1)\sqrt{T}}}{(d^{2\sqrt{T}}-1)^2} \right] = 0$$

and

$$\lim_{T\to\infty} \left[\sum_{k=1}^{N-1} \frac{2k-1}{1 + d^{\sqrt{T}[k-(\nu+1/3)]}} \right] = (N-1)^2.$$

Therefore

$$\begin{aligned}
\lim_{T\to\infty} \mathrm{E}\left[\mathrm{X}_T^{(\mathrm{Map})}(1)^2 \,\middle|\, X_T \right] &= (\lceil \nu + 1/3 \rceil - 1)^2 \\
&= (\lceil \nu - 2/3 + 1 \rceil - 1)^2 \\
&= (\lceil \nu - 2/3 \rceil)^2 \\
&= \left(\left\lceil \alpha X_T + \lambda - \frac{2}{3} \right\rceil \right)^2.
\end{aligned}$$

□

Proposition 12. *Let $\widehat{\alpha}$ and $\widehat{\lambda}$ be estimators of α and λ such that*

$$\sqrt{T} \begin{pmatrix} \widehat{\alpha} - \alpha \\ \widehat{\lambda} - \lambda \end{pmatrix} \xrightarrow{D} \mathcal{N}_2(\mathbf{0}, \Sigma).$$

Then, given X_T, the sequence of estimators $\left(X_T^{(\mathrm{Map})}(1)\right)_{T\geqslant 1}$ is weakly conditionally consistent for δ, this is, for each $\epsilon > 0$

$$\lim_{T\longrightarrow\infty} \mathrm{P}\left(\left|X_T^{(\mathrm{Map})}(1)-\delta\right| \geqslant \epsilon \,\middle|\, X_T\right) = 0,$$

where $\delta = [\alpha X_T + \lambda - 2/3]$.

Proof. Note that Chebyshev's Inequality is also true for a random variable conditioned to a random event, i. e., if Y is a random variable, A is an event, then, for each $\epsilon > 0$, we have

$$\mathrm{P}(\,Y \geqslant \epsilon \,|\, A\,) \leqslant \frac{\mathrm{E}[\,Y^2 \,|\, A\,]}{\epsilon^2}. \tag{4.20}$$

In fact,

$$\begin{aligned}
\mathrm{E}[\,Y^2|A\,] &= \mathrm{E}\left[\,I_{\{Y\geqslant\epsilon\}}Y^2 \,|\, A\,\right] + \underbrace{\mathrm{E}\left[\,I_{\{Y<\epsilon\}}Y^2 \,|\, A\,\right]}_{\geqslant 0}\\
&\geqslant \mathrm{E}\left[\,I_{\{Y\geqslant\epsilon\}}Y^2 \,|\, A\,\right]\\
&\geqslant \epsilon^2\,\mathrm{E}\left[\,I_{\{Y\geqslant\epsilon\}} \,|\, A\,\right]\\
&= \epsilon^2\,\mathrm{P}\left(Y \geqslant \epsilon \,|\, A\,\right),
\end{aligned}$$

where I_B represents the indicator function of the event B.
Then if we consider $Y = \left|X_T^{(\mathrm{Map})}(1) - \delta\right|$ and the event $A = [X_T = x_T]$, from Equation (4.20) we obtain

$$\mathrm{P}\left(\,\left|X_T^{(\mathrm{Map})}(1)-\delta\right| \geqslant \epsilon \,\middle|\, X_T\right) \leqslant \frac{\mathrm{E}\left[\,\left(X_T^{(\mathrm{Map})}(1)-\delta\right)^2 \,\middle|\, X_T\right]}{\epsilon^2}, \tag{4.21}$$

then, using Proposition 10 and Proposition 11 we obtain

$$\begin{aligned}
\lim_{T\longrightarrow\infty} \mathrm{P}\left(\,\left|X_T^{(\mathrm{Map})}(1)-\delta\right| \geqslant \epsilon \,\middle|\, X_T\right) &\leqslant \lim_{T\longrightarrow\infty} \frac{\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)^2 \,\middle|\, X_T\right] - 2\delta\mathrm{E}\left[X_T^{(\mathrm{Map})}(1) \,\middle|\, X_T\right]}{\epsilon^2}\\
&\quad + \frac{\mathrm{E}\left[\delta^2 \,\middle|\, X_T\right]}{\epsilon^2}\\
&= \lim_{T\longrightarrow\infty} \frac{\mathrm{Var}\left[X_T^{(\mathrm{Map})}(1)^2 \middle| X_T\right] + \left\{\mathrm{E}\left[X_T^{(\mathrm{Map})}(1)\middle| X_T\right]\right\}^2}{\epsilon^2}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{-2\delta \mathrm{E}\left[X_T^{(\mathrm{Map})}(1)\Big|X_T\right]+\delta^2}{\epsilon^2}\\
&=\frac{0+\delta^2-2\delta^2+\delta^2}{\epsilon^2}\\
&=0.
\end{aligned}
$$

Considering $\epsilon < 1$, we have $\lim_{T\longrightarrow\infty} \mathrm{P}\Big(X_T^{(\mathrm{Map})}(1) = \delta \mid X_T\Big) = 1$ □

Proposition 13. *Let $\widehat{\alpha}$ and $\widehat{\lambda}$ be estimators of α and λ such that*

$$
\sqrt{T}\begin{pmatrix}\widehat{\alpha}-\alpha\\ \widehat{\lambda}-\lambda\end{pmatrix}\xrightarrow{D}\mathcal{N}_2(\mathbf{0},\Sigma).
$$

Then the sequence of estimators $\big(X_T^{(\mathrm{Map})}(1)\big)_{T\geqslant 1}$ is weakly consistent for δ, this is, for each $\epsilon > 0$

$$
\lim_{T\longrightarrow\infty} \mathrm{P}\big(\big|X_T^{(\mathrm{Map})}(1)-\delta\big|\geqslant \epsilon\,\big)=0, \tag{4.22}
$$

where $\delta = [\alpha X_T + \lambda - 2/3]$.

Proof. Note that we can write $\mathrm{P}\big(\big|X_T^{(\mathrm{Map})}(1)-\delta\big|\geqslant \epsilon\,\big)$ as follows

$$
\begin{aligned}
\mathrm{P}\big(\big|X_T^{(\mathrm{Map})}(1)-\delta\big|\geqslant \epsilon\,\big) &= \mathrm{E}\Big\{I_{\left[\mathrm{P}\left(\left|X_T^{(\mathrm{Map})}(1)-\delta\right|\geqslant\epsilon\,\right)\right]}\Big\}\\
&=\mathrm{E}\left[\mathrm{E}\left\{I_{\left[\mathrm{P}\left(\left|X_T^{(\mathrm{Map})}(1)-\delta\right|\geqslant\epsilon\,\right)\right]}\,\middle|\,X_T\right\}\right]\\
&=\mathrm{E}\left[\mathrm{P}\left(\big|X_T^{(\mathrm{Map})}(1)-\delta\big|\geqslant\epsilon\,\middle|\,X_T\right)\right].
\end{aligned}
$$

Then, given that the sequence of random variables $\left(\mathrm{P}\left(\big|X_T^{(\mathrm{Map})}(1)-\delta\big|\geqslant\epsilon^2\,\middle|\,X_T\right)\right)_{T\geqslant 1}$ is dominated by the constant random variable 1, i. e., $\mathrm{P}\left(\big|X_T^{(\mathrm{Map})}(1)-\delta\big|\geqslant\epsilon\,\middle|\,X_T\right)\leqslant 1$, we can use the dominated convergence theorem and from the last proposition we obtain

$$
\begin{aligned}
\lim_{T\longrightarrow\infty} \mathrm{P}\big(\big|X_T^{(\mathrm{Map})}(1)-\delta\big|\geqslant \epsilon\,\big) &= \lim_{T\longrightarrow\infty}\mathrm{E}\left[\mathrm{P}\left(\big|X_T^{(\mathrm{Map})}(1)-\delta\big|\geqslant\epsilon\,\middle|\,X_T\right)\right]\\
&=\mathrm{E}\left[\lim_{T\longrightarrow\infty}\mathrm{P}\left(\big|X_T^{(\mathrm{Map})}(1)-\delta\big|\geqslant\epsilon\,\middle|\,X_T\right)\right]\\
&=0.
\end{aligned}
$$

□

Given that $X_T^{(\mathrm{Map})}(1)$ and $\delta = \lceil \alpha X_T + \lambda - 2/3 \rceil$ are non-negative integer random variables and Equation (4.22) is true for all $\epsilon > 0$, if we consider $\epsilon < 1$ we will have

$$\mathrm{P}\big(\big|X_T^{(\mathrm{Map})}(1) - \delta\big| \geqslant \epsilon \big) = \mathrm{P}\big(X_T^{(\mathrm{Map})}(1) \neq \delta\big).$$

Therefore, Equation (4.22) of Proposition 13 can be expressed as

$$\lim_{T\to\infty} \mathrm{P}\left(X_T^{(\mathrm{Map})}(1) \neq \left\lceil \alpha X_T + \lambda - \frac{2}{3} \right\rceil \right) = 0.$$

Proposition 14. *Let $\widehat{\alpha}$ and $\widehat{\lambda}$ be estimators of α and λ such that*

$$\sqrt{T} \begin{pmatrix} \widehat{\alpha} - \alpha \\ \widehat{\lambda} - \lambda \end{pmatrix} \xrightarrow{D} \mathcal{N}_2(\mathbf{0}, \Sigma).$$

Then given X_T, the sequence of estimators $\big(X_T^{(\mathrm{Map})}(1)\big)_{T\geqslant 1}$ is strongly conditionally consistent for δ, this is

$$\mathrm{P}\Big(\lim_{T\longrightarrow\infty} X_T^{(\mathrm{Map})}(1) = \delta \,\Big|\, X_T \Big) = 1,$$

where $\delta = \lceil \alpha X_T + \lambda - 2/3 \rceil$.

Proof. From Equation (4.21) we have

$$\mathrm{P}\Big(\big|X_T^{(\mathrm{Map})}(1) - \delta\big| \geqslant \epsilon \,\Big|\, X_T \Big) \leqslant \frac{\mathrm{E}\Big[\big|X_T^{(\mathrm{Map})}(1) - \delta\big| \,\Big|\, X_T \Big]}{\epsilon},$$

then,

$$\begin{aligned} \sum_{T=1}^{\infty} \mathrm{P}\Big(\big|X_T^{(\mathrm{Map})}(1) - \delta\big| \geqslant \epsilon \,\Big|\, X_T \Big) &\leqslant \frac{1}{\epsilon} \sum_{T=1}^{\infty} \sum_{j=1}^{\infty} |\, j - \delta|\, \mathrm{P}\big(X_T^{(\mathrm{Map})}(1) = j \,\big|\, X_T\big) \\ &= \frac{1}{\epsilon} \Bigg[\sum_{T=1}^{\infty} \sum_{j=1}^{\delta-1} (\delta - j)\, \mathrm{P}\big(X_T^{(\mathrm{Map})}(1) = j \,\big|\, X_T\big) \\ &\qquad + \sum_{T=1}^{\infty} \sum_{j=\delta}^{\infty} (\, j - \delta)\, \mathrm{P}\big(X_T^{(\mathrm{Map})}(1) = j \,\big|\, X_T\big) \Bigg] \\ &= \frac{1}{\epsilon} \Bigg[\sum_{j=1}^{\delta-1} (\delta - j) \underbrace{\sum_{T=1}^{\infty} \mathrm{P}\big(X_T^{(\mathrm{Map})}(1) = j \,\big|\, X_T\big)}_{B_1} \\ &\qquad + \underbrace{\sum_{j=1}^{\infty} j \sum_{T=1}^{\infty} \mathrm{P}\big(X_T^{(\mathrm{Map})}(1) = j + \delta \,\big|\, X_T\big)}_{B_2} \Bigg]. \end{aligned} \tag{4.23}$$

Our aim is to prove that

$$\sum_{T=1}^{\infty} \mathrm{P}\Big(\big|X_T^{(\mathrm{Map})}(1) - \delta\big| \geqslant \epsilon \,\Big|\, X_T \Big) < \infty,$$

and use the Borel-Cantelli lemma, so, we need to prove that $B_1 < \infty$ and $B_2 < \infty$. Note that,

$$\begin{aligned}
\mathrm{P}\big(X_T^{(\mathrm{Map})}(1) = j \,\big|\, X_T\big) &= \mathrm{P}\Big(\Big\lceil \widehat{\alpha} X_T + \widehat{\lambda} - \frac{2}{3} \Big\rceil = j \,\Big|\, X_T \Big) \\
&= \mathrm{P}\Big(j - 1 < \widehat{\alpha} X_T + \widehat{\lambda} - \frac{2}{3} \leqslant j \,\Big|\, X_T \Big) \\
&= \mathrm{P}\Big(\widehat{\alpha} X_T + \widehat{\lambda} \leqslant j + \frac{2}{3} \,\Big|\, X_T \Big) - \mathrm{P}\Big(\widehat{\alpha} X_T + \widehat{\lambda} \leqslant j - \frac{1}{3} \,\Big|\, X_T \Big) \\
&\approx \mathrm{P}\left(\mathcal{Z} \leqslant \frac{\sqrt{T}(j + \frac{2}{3} - \nu)}{\sigma_{\alpha,\lambda}} \,\middle|\, X_T \right) - \mathrm{P}\left(\mathcal{Z} \leqslant \frac{\sqrt{T}(j - \frac{1}{3} - \nu)}{\sigma_{\alpha,\lambda}} \,\middle|\, X_T \right) \\
&= \mathrm{P}\big(\mathcal{Z} \leqslant c_{j+1}\sqrt{T}\big) - \mathrm{P}\big(\mathcal{Z} \leqslant c_j\sqrt{T}\big) \\
&= \Phi\left(c_{j+1}\sqrt{T}\right) - \Phi\left(c_j\sqrt{T}\right) \\
&= \big[1 - \Phi\left(-c_j\sqrt{T}\right)\big] - \big[1 - \Phi\left(-c_{j+1}\sqrt{T}\right)\big] \qquad (4.24)
\end{aligned}$$

where $\mathcal{Z}$ is a random variable with standard normal distribution, $\Phi(\cdot)$ denotes the cumulative distribution function of $\mathcal{Z}$ and $c_j = (j - 1/3 - \nu)/\sigma_{\alpha,\lambda}$ and j is a fixed integer, such as $1 \leqslant j \leqslant \delta - 1$.

Note that $c_j < 0$ if $j < \delta$. In fact,

$$j < \delta \quad \Longrightarrow \quad j - 1/3 - \nu < \delta - 1/3 - \nu,$$

given that $\delta = \lceil \nu - 2/3 \rceil$ we have

$$j - 1/3 - \nu < \lceil \nu - 2/3 \rceil - (\nu - 2/3) - 1,$$

where $\Big[\lceil \nu - 2/3 \rceil - (\nu - 2/3)\Big] \in (0,1) \quad \Longrightarrow \quad c_j = (j - 1/3 - \nu)/\sigma_{\alpha,\lambda} < 0.$

On the other hand,

$$c_j = 0 \quad \Longleftrightarrow \quad j = 1/3 + \nu$$

$$\iff \quad \delta = \lceil j - 1/3 - 2/3 \rceil = j - 1 \quad \iff \quad j = \delta + 1.$$

Therefore c_j is not zero for $j < \delta$.

By Zelen and Severo [1964] if $x > 0$ we can approximate $\mathrm{P}(\mathcal{Z} > x)$ by

$$1 - \Phi(x) \approx m_1 \exp(-x^2/2) \sum_{i=1}^{3} \frac{(-1)^{i+1} m_{i+1}}{(1 + m_5 x)^i},$$

where $m_1 = \sqrt{1/2\pi}$, $m_2 = 0.4361836$, $m_3 = 0.1201676$, $m_4 = 0.9372980$ and $m_5 = 0.33267$.

Therefore we approximate $\mathrm{P}(\mathcal{Z} > -c_j\sqrt{T})$ in Equation (4.24) by

$$\sum_{T=1}^{\infty} \mathrm{P}(\mathcal{Z} > -c_j\sqrt{T}) \approx m_1 \sum_{i=1}^{3} (-1)^{i+1} m_{i+1} \sum_{T=1}^{\infty} \frac{\exp(-c_j^2\, T/2)}{(1 - m_5 c_j \sqrt{T})^i}. \tag{4.25}$$

Notice that for each i (i = 1, 2, 3), from the Ratio Test, we have

$$\sum_{T=1}^{\infty} \frac{\exp(-c_j^2\, T/2)}{(1 + m_5 c_j \sqrt{T})^i} < \infty.$$

In fact, if $a_{T,i} = \dfrac{\exp(-c_j^2\, T/2)}{(1 - m_5 c_j \sqrt{T})^i}$, then

$$\lim_{T \longrightarrow \infty} \frac{a_{T+1,i}}{a_{T,i}} = \lim_{T \longrightarrow \infty} \left(\frac{1 - m_5 c_j \sqrt{T}}{1 - m_5 c_j \sqrt{T+1}} \right)^i \exp(-c_j^2/2) < 1.$$

The last expression is true if we suppose $1/3 + \nu \neq j$, in other words, $1/3 + \widehat{\alpha} X_T + \widehat{\lambda} \neq j$ with $1 \leqslant j \leqslant \delta - 1$.

Also, Equation (4.25) holds if we replace c_j by c_{j+1}, thus we have proved that $B_1 = \sum_{T=1}^{\infty} \mathrm{P}\big(X_T^{(\mathrm{Map})}(1) = j \,\big|\, X_T \big) < \infty$ for $1 \leqslant j \leqslant \delta - 1$.

Besides for fixed j, $j \geqslant 1$, we can approximate $\mathrm{P}\big(X_T^{(\mathrm{Map})}(1) = j + \delta \,\big|\, X_T \big)$ by

$$\mathrm{P}\big(X_T^{(\mathrm{Map})}(1) = j + \delta \,\big|\, X_T \big) \approx \mathrm{P}(\mathcal{Z} > c_{j+\delta}\sqrt{T}) - \mathrm{P}(\mathcal{Z} > c_{j+\delta+1}\sqrt{T}).$$

Notice that if $\nu - 2/3$ is not an integer, then, $c_{j+\delta} > 0, \forall j \geqslant 1$. In fact,

$$\sigma_{\alpha,\lambda} c_{j+\delta} = j + \delta - \nu - 1/3$$

$$= \delta - \big(\nu - 2/3\big) + j - 1$$
$$= \underbrace{\lceil \nu - 2/3 \rceil - \big(\nu - 2/3\big)}_{>0 \text{ if } (\nu-2/3)\notin\mathbb{Z}} + j - 1.$$

Thus,

$$\frac{1}{\big(1 + m_5 c_{j+\delta}\sqrt{T}\big)^i} < 1, \tag{4.26}$$

and using the Zelen and Severo [1964] normal approximation again we have

$$\begin{aligned}
\sum_{j=1}^{\infty} j \sum_{T=1}^{\infty} \mathrm{P}\big(\mathcal{Z} > c_{j+\delta}\sqrt{T}\big) &\approx m_1 \sum_{i=1}^{3} (-1)^{i+1} m_{i+1} \sum_{j=1}^{\infty} j \sum_{T=1}^{\infty} \frac{\exp(-c_{j+\delta}^2\, T/2)}{(1 + m_5 c_{j+\delta}\sqrt{T})^i} \\
&< m_1 \sum_{i=1}^{3} (-1)^{i+1} m_{i+1} \sum_{j=1}^{\infty} j \sum_{T=1}^{\infty} \big[\exp(-c_{j+\delta}^2/2)\big]^T && (4.27) \\
&= m_1 \sum_{i=1}^{3} (-1)^{i+1} m_{i+1} \sum_{j=1}^{\infty} \frac{j}{\exp(c_{j+\delta}^2/2) - 1} && (4.28) \\
&< \infty && (4.29)
\end{aligned}$$

Equation (4.27) is obtained by using Equation (4.26). Equation (4.28) is true since $\exp(-c_{j+\delta}^2/2) < 1$ and

$$\sum_{T=1}^{\infty} \Big[\exp(-c_{j+\delta}^2/2)\Big]^T = \frac{1}{\exp(c_{j+\delta}^2/2) - 1}.$$

Equation (4.29) holds since $\displaystyle\sum_{j=1}^{\infty} \frac{j}{\exp(c_{j+\delta}^2/2) - 1}$ converges. In fact, if

$$d_j = \frac{j}{\exp(c_{j+\delta}^2/2) - 1},$$

then

$$\frac{d_{j+1}}{d_j} = \left(\frac{j+1}{j}\right)\left(\frac{\exp(c_{j+\delta}^2/2) - 1}{\exp(c_{j+\delta+1}^2/2) - 1}\right).$$

We have

$$\lim_{j\longrightarrow\infty} \left[\frac{\exp(c_{j+\delta}^2/2) - 1}{\exp(c_{j+\delta+1}^2/2) - 1}\right] = \lim_{j\longrightarrow\infty} \left[\frac{\exp(c_{j+\delta}^2/2) c_{j+\delta} (1/\sigma_{\alpha,\lambda})}{\exp(c_{j+\delta+1}^2/2) c_{j+\delta+1} (1/\sigma_{\alpha,\lambda})}\right] \tag{4.30}$$

$$
\begin{aligned}
&= \lim_{j \longrightarrow \infty} \left[\frac{c_{j+\delta}}{c_{j+\delta+1}} \right] \left[\frac{\exp(c^2_{j+\delta}/2)}{\exp(c^2_{j+\delta+1}/2)} \right] \\
&= 0,
\end{aligned}
$$

Equation (4.30) is obtained by applaying L'Hôpital rule with respect to j.
Given that

$$
\begin{aligned}
\lim_{j \longrightarrow \infty} \left[\frac{c_{j+\delta}}{c_{j+\delta+1}} \right] &= \lim_{j \longrightarrow \infty} \frac{\left(j+\delta-\frac{1}{3}-\nu\right)/\sigma_{\alpha,\lambda}}{\left(j+\delta+\frac{2}{3}-\nu\right)/\sigma_{\alpha,\lambda}} \\
&= 1.
\end{aligned}
$$

and

$$
\begin{aligned}
\lim_{j \longrightarrow \infty} \left[\frac{\exp(c^2_{j+\delta}/2)}{\exp(c^2_{j+\delta+1}/2)} \right] &= \lim_{j \longrightarrow \infty} \exp\left\{ \frac{1}{2\sigma^2_{\alpha,\lambda}} \left[\left(j+(\delta-1/3-\nu)\right)^2 \right.\right. \\
&\qquad \left.\left. - \left(j+(\delta+2/3-\nu)\right)^2 \right] \right\} \\
&= \lim_{j \longrightarrow \infty} \exp\Big[j^2 + 2j(\delta-1/3-\nu) + (\delta-1/3-\nu)^2 \\
&\qquad - j^2 - 2j(\delta+2/3-\nu) - (\delta+2/3-\nu)^2 \Big] \Big/ 2\sigma^2_{\alpha,\lambda} \\
&= \exp\left(\frac{-2\delta+2\nu-1/3}{2\sigma^2_{\alpha,\lambda}} \right) \lim_{j \longrightarrow \infty} \left[\exp\left(-j/\sigma^2_{\alpha,\lambda} \right) \right] \\
&= 0,
\end{aligned}
$$

Therefore by using the Ratio Test we obtain

$$
\begin{aligned}
\lim_{j \longrightarrow \infty} \frac{d_{j+1}}{d_j} &= \lim_{j \longrightarrow \infty} \left[\frac{j+1}{j} \right] \lim_{j \longrightarrow \infty} \left[\frac{\exp(c^2_{j+\delta}/2)-1}{\exp(c^2_{j+\delta+1}/2)-1} \right] \\
&= 1 \cdot 0 = 0 < 1.
\end{aligned}
$$

We have proved that

$$
B_2 = \sum_{j=1}^{\infty} j \sum_{T=1}^{\infty} \mathrm{P}\left(\mathcal{Z} > c_{j+\delta+1}\sqrt{T} \right) < \infty. \tag{4.31}
$$

Thus from Equation (4.23) and using (4.29) and Equation (4.31) we have proved

$$
\sum_{T=1}^{\infty} \mathrm{P}\left(\left| X_T^{(\mathrm{Map})}(1) - \delta \right| \geqslant \epsilon \,\middle|\, X_T \right) < \infty.
$$

Therefore from Borel-Cantelli lemma we have that

$$P\Big(\Big[\big| X_T^{(\mathrm{Map})}(1) - \delta \big| \geqslant \epsilon \;\; i.\ o. \;\Big|\; X_T \Big] \Big) = 0,$$

and, this is equivalent to

$$P\Big(\lim_{T \longrightarrow \infty} X_T^{(\mathrm{Map})}(1) = \delta \,\Big|\, X_T \Big) = 1.$$

$\square$

Proposition 15. *Let $\widehat{\alpha}$ and $\widehat{\lambda}$ be estimators of α and λ such that*

$$\sqrt{T} \begin{pmatrix} \widehat{\alpha} - \alpha \\ \widehat{\lambda} - \lambda \end{pmatrix} \xrightarrow{D} \mathcal{N}_2(\mathbf{0}, \Sigma).$$

Then given X_T, the sequence of estimators $\big(X_T^{(\mathrm{Map})}(1)\big)_{T \geqslant 1}$ is strongly consistent for δ, this is

$$P\Big(\lim_{T \longrightarrow \infty} X_T^{(\mathrm{Map})}(1) = \delta \Big) = 1,$$

where $\delta = [\alpha X_T + \lambda - 2/3]$.

Proof. Note that we can write $P\Big(\lim_{T \longrightarrow \infty} X_T^{(\mathrm{Map})}(1) = \delta \Big)$ as follows

$$\begin{aligned}
P\Big(\lim_{T \longrightarrow \infty} X_T^{(\mathrm{Map})}(1) = \delta \Big) &= E\Big\{ I_{\big[P\big(\lim_{T \longrightarrow \infty} X_T^{(\mathrm{Map})}(1) = \delta \big)\big]} \Big\} \\
&= E\Big[E\Big\{ I_{\big[P\big(\lim_{T \longrightarrow \infty} X_T^{(\mathrm{Map})}(1) = \delta \big)\big]} \,\Big|\, X_T \Big\} \Big] \\
&= E\Big[P\Big(\lim_{T \longrightarrow \infty} X_T^{(\mathrm{Map})}(1) = \delta \,\Big|\, X_T \Big) \Big] \qquad (4.32) \\
&= E[1] = 1,
\end{aligned}$$

Equation (4.32) is true from Proposition 14. $\square$

4.5 Forecasting two and $h-$steps ahead

Now we focus on the forecast two-steps ahead. Note that we can obtain the probability function of $X_{T+2} \mid X_T$ using the PGF; thus from equation (4.10) for $h = 2$, we have

$$\begin{aligned} \mathcal{G}_{X_{T+2}|X_T}(s) &= \exp\left[\frac{\lambda\, g\big(\log(s)\big) + \nu\, g\big(g(\log(s))\big)}{\alpha}\right] \\ &= \exp\Big\{\lambda(s-1) + \nu\,\big[\exp(\alpha(s-1)) - 1\big]\Big\}, \end{aligned} \tag{4.33}$$

where $\nu = \alpha X_T + \lambda$. Remember that from the PGF we can obtain the probability function using the relation

$$\mathrm{P}\big(X_{T+2} = k \,\big|\, X_T\big) = \frac{1}{k!}\left\{\left.\frac{d^{(k)}\left[\mathcal{G}_{X_{T+2}|X_T}(s)\right]}{ds^k}\right|_{s=0}\right\},$$

where $\dfrac{d^{(k)}[h(x)]}{dx}$ denotes the $k-$th derivative of function $h(x)$. The conditional probability function of X_{T+2} given X_T is also given in the following proposition.

Proposition 16. *The probability function of $X_{T+2} \mid X_T$ is*

$$\mathrm{P}\big(X_{T+2} = k \,\big|\, X_T\big) = \lambda^k\, \mathcal{C}_{\nu,\alpha,\lambda} \sum_{j=0}^{k} \sum_{i=0}^{j} \left(\frac{\alpha}{\lambda}\right)^j \frac{S(j,i)\,[\,\nu\,\exp(-\alpha)\,]^{\,i}}{j\,!\,(k-j)!},$$

where $\nu = \alpha X_T + \lambda$, $\mathcal{C}_{\nu,\alpha,\lambda} = \exp\big[\nu(\exp(-\alpha) - 1) - \lambda\big]$ and $S(j,i)$ represents the Stirling number of the second kind.

Proof.

$$\begin{aligned} \mathrm{P}\big(X_{T+2} = k \,\big|\, X_T = x_T\big) &= \sum_{n=0}^{\infty} \mathrm{P}\big(X_{T+2} = k \,\big|\, X_{T+1} = n, X_T = x_T\big)\mathrm{P}\big(X_{T+1} = n \,\big|\, X_T = x_T\big) \\ &= \sum_{n=0}^{\infty} \mathrm{P}\big(X_{T+2} = k \,\big|\, X_{T+1} = n\big)\mathrm{P}\big(X_{T+1} = n \,\big|\, X_T = x_T\big) \\ &= \sum_{n=0}^{\infty} \left[\frac{\exp[-(\alpha n + \lambda)](\alpha n + \lambda)^k}{k!}\right]\left[\frac{\exp[-(\alpha x_T + \lambda)](\alpha x_T + \lambda)^n}{n!}\right] \\ &= \frac{\exp(-\nu - \lambda)}{k!} \sum_{n=0}^{\infty} \frac{[\exp(-\alpha)\nu]^n}{n!} \sum_{j=0}^{k} \binom{k}{j} (\alpha n)^j \lambda^{k-j} \end{aligned}$$

$$= \frac{\lambda^k}{k!} \exp\left[-\nu - \lambda + \nu \exp(-\alpha)\right] \sum_{j=0}^{k} \binom{k}{j} \left(\frac{\alpha}{\lambda}\right)^j \times$$

$$\sum_{n=0}^{\infty} n^j \frac{[\exp(-\nu \exp(-\alpha))][\nu \exp(-\alpha)]^n}{n!}$$

$$= \frac{\lambda^k}{k!} \exp\left[-\nu - \lambda + \nu \exp(-\alpha)\right] \sum_{j=0}^{k} \binom{k}{j} \left(\frac{\alpha}{\lambda}\right)^j \mathrm{E}\left[\mathcal{W}^j\right] \tag{4.34}$$

$$= \lambda^k \, \mathcal{C}_{\nu,\alpha,\lambda} \sum_{j=0}^{k} \sum_{i=0}^{j} \left(\frac{\alpha}{\lambda}\right)^j \frac{S(j,i)[\nu \exp(-\alpha)]^i}{j!\,(k-j)!}, \tag{4.35}$$

where $\mathcal{W} \sim \mathrm{Po}\big(\nu \exp(-\alpha)\big)$ in equation (4.34) and equation (4.35) is obtained from the expression for the $j-$th ordinary moment of Poisson distribution given in equation (4.4). $\square$

In particular, note that $\mathrm{P}\big(X_{T+2} = 0 \,\big|\, X_T\big) = \mathcal{C}_{\nu,\alpha,\lambda}$ can be obtained from the last proposition using that $S(0,0) = 1$ or from PGF two-steps ahead given in equation (4.33) taking $s = 0$. Note that Proposition 4.33 provides an explicit expression for the distribution of $X_{T+2} \,|\, X_T$. Then we can use the conditional median and conditional mode to predict X_{T+2}. Therefore, given that $X_1, X_2, \ldots, X_T$ are known we can predict X_{T+2} by using

$$\begin{cases} X_T^{(\mathrm{Mn})}(2), & \text{the conditional median of } X_{T+2} \,|\, X_T \\ X_T^{(\mathrm{Md})}(2), & \text{the conditional mode of } X_{T+2} \,|\, X_T \\ X_T^{(\mathrm{Ei})}(2), & \text{the integer part of } \mathrm{E}\big[X_{T+2} \,|\, X_T\big], \end{cases} \tag{4.36}$$

where $X_T^{(\mathrm{Ei})}(2)$ is obtained from Proposition 6 replacing $h = 2$, the explicit expression of $X_T^{(\mathrm{Ei})}(2)$ is

$$X_T^{(\mathrm{Ei})}(2) = \big\lfloor \widehat{\alpha}^2 X_T + (1 + \widehat{\alpha})\widehat{\lambda} \big\rfloor,$$

where $\widehat{\alpha}$ and $\widehat{\lambda}$ represent estimators of α and λ, respectively.

Since we have not gotten an explicit expression for $\mathrm{P}\big(X_{T+h} = k \,\big|\, X_T\big)$ for $h \geqslant 2$ yet, we propose to use recursively the forecast $h-1$ steps ahead to find the forecast h steps ahead as follows: given that $X_1, X_2, \ldots, X_T$ are known, we propose

$$\begin{aligned} &\text{to predict } X_{T+1} \text{ by using } X_T^{(\mathrm{Map})}(1) = \left\lceil \widehat{\alpha} X_T + \widehat{\lambda} - \frac{2}{3} \right\rceil; \\ &\text{to predict } X_{T+h} \text{ by using } X_T^{(\mathrm{Map})}(h) = \left\lceil \widehat{\alpha} X_T^{(\mathrm{Map})}(h-1) + \widehat{\lambda} - \frac{2}{3} \right\rceil \text{ for } h \geqslant 2, \end{aligned} \tag{4.37}$$

where $\widehat{\alpha}$ and $\widehat{\lambda}$ represent estimators of α and λ respectively.
Finally since we provide an analytic expression for $\mathrm{E}\left[X_{T+h}\,|\,X_T\right]$ in Proposition 6, we also propose to predict X_{T+h} by using

$$X_T^{(\mathrm{Ei})}(h) = \left\lfloor \widehat{\alpha}^h X_T + \left(\frac{1-\widehat{\alpha}^h}{1-\widehat{\alpha}}\right)\widehat{\lambda} \right\rfloor.$$

Note that if $\widehat{\alpha}X_T + \widehat{\lambda}$ is non integer then $X_T^{(\mathrm{Ei})}(1) = X_T^{(\mathrm{Md})}(1) = \left\lfloor \widehat{\alpha}X_T + \widehat{\lambda}\right\rfloor$.

4.6 One-step ahead prediction interval

Given that $X_1, X_2, \ldots, X_T$ are known, we can construct the one-sided or the two-sided $100(1-\gamma)\%$ prediction intervals for X_{T+1}. First, consider the construction of the two-sided $100(1-\gamma)\%$ prediction interval for X_{T+1}

1. Find $\widehat{\alpha}$ and $\widehat{\lambda}$, so the estimated distribution of $X_{T+1}\,|\,X_T$ is the Poisson distribution with mean $\widehat{\mu}_{T+1} = \widehat{\alpha}X_T + \widehat{\lambda}$
2. Find the greatest integer value l_{T+1} such that
$$\mathrm{P}\left(W_{\widehat{\mu}_{T+1}} < l_{T+1}\right) \leqslant \frac{\gamma}{2},$$
where $W_{\widehat{\mu}_{T+1}}$ represent a random variable Poisson distributed with mean $\widehat{\mu}_{T+1}$
3. Find the lowest integer value u_{T+1} such that
$$\mathrm{P}\left(W_{\widehat{\mu}_{T+1}} \leqslant u_{T+1}\right) \geqslant 1-\frac{\gamma}{2}.$$

Then,

$$\begin{aligned}\mathrm{P}\left(l_{T+1} \leqslant W_{\widehat{\mu}_{T+1}} \leqslant u_{T+1}\right) &= \mathrm{P}\left(W_{\widehat{\mu}_{T+1}} \leqslant u_{T+1}\right) - \mathrm{P}\left(W_{\widehat{\mu}_{T+1}} < l_{T+1}\right)\\ &\geqslant 1-\frac{\gamma}{2}-\frac{\gamma}{2}\\ &= 1-\gamma,\end{aligned}$$

therefore, $\left[\, l_{T+1},\, u_{T+1}\right]$ is a prediction interval for X_{T+1} with confidence of least $100(1-\gamma)\%$. Besides, notice that the length $u_{T+1} - l_{T+1}$ is minimal.

Now consider the construction of the upper $100(1-\gamma)\%$ prediction interval for X_{T+1}

1. Find $\widehat{\alpha}$ and $\widehat{\lambda}$, then given X_T, the estimated distribution of $X_{T+1} \mid X_T$ is the Poisson distribution with mean $\widehat{\mu}_{T+1} = \widehat{\alpha} X_T + \widehat{\lambda}$

2. Find the lowest integer value u_{T+1} such that

$$\mathrm{P}\big(W_{\widehat{\mu}_{T+1}} \leqslant u_{T+1} \big) \geqslant 1 - \gamma.$$

Then,

$$\begin{aligned}
\mathrm{P}\left(0 \leqslant W_{\widehat{\mu}_{T+1}} \leqslant u_{T+1} \right) &= \mathrm{P}\left(W_{\widehat{\mu}_{T+1}} \leqslant u_{T+1} \right) - \mathrm{P}\left(W_{\widehat{\mu}_{T+1}} < 0 \right) \\
&\geqslant 1 - \gamma - 0 \\
&= 1 - \gamma,
\end{aligned}$$

therefore, $\left[0,\, u_{T+1} \right]$ is a smaller upper $100(1-\gamma)\%$ prediction interval for X_{T+1}.

Finally, the lower $100(1-\gamma)\%$ prediction interval for X_{T+1} is found similarly to the construction of the upper prediction interval

1. Find $\widehat{\alpha}$ and $\widehat{\lambda}$, then given X_T, the estimated distribution of $X_{T+1} \mid X_T$ is the Poisson distribution with mean $\widehat{\mu}_{T+1} = \widehat{\alpha} X_T + \widehat{\lambda}$

2. Find the greatest integer value l_{T+1} such that

$$\mathrm{P}\big(W_{\widehat{\mu}_{T+1}} < l_{T+1} \big) \leqslant \gamma.$$

Then,

$$\begin{aligned}
\mathrm{P}\left(l_{T+1} \leqslant W_{\widehat{\mu}_{T+1}} \right) &= 1 - \mathrm{P}\left(W_{\widehat{\mu}_{T+1}} < l_{T+1} \right) \\
&\geqslant 1 - \gamma,
\end{aligned}$$

therefore, $\left[l_{T+1},\, \infty \right)$ is a smaller lower $100(1-\gamma)\%$ prediction interval for X_{T+1}.

4.7 Monte Carlo results for forecasting one-step ahead

In this section we present a Monte Carlo study that investigates and compares the performances of the forecasts proposed in the last section: exact median, approximate median and mode of the conditional distribution. Using the RMSE and the MAE we compared them in two scenarios, considering known and unknown parameters. For the simulation study, we used 10000 Monte Carlo samples, and we considered different values of α and λ, namely, $\alpha = 0.1$, 0.5, 0.9 and $\lambda = 0.5$, 1, 3, 5. Finally, we considered the sample sizes $T = 25$, 50 and 100.

In the first scenario we consider known parameters. Since we do not know the marginal stationary distribution, we generate N additional values to generate x_1. For $r = 1, \ldots,$ 10000 and $N = 1500$ we generate

1. $y_1^{(r)} = 0$
2. $y_t^{(r)} = \text{Po}(\alpha y_{t-1}^{(r)} + \lambda), \quad t = 2, \ldots, N+1 \quad \dashrightarrow \quad y_1^{(r)}, y_2^{(r)}, \ldots, y_{N+1}^{(r)}$

   $\dashrightarrow \quad x_1^{(r)} = y_{N+1}^{(r)}$
3. $x_1^{(r)} \quad \dashrightarrow \quad x_2^{(r)} = \text{Po}(\alpha x_1^{(r)} + \lambda)$

Table 4.1 shows the RMSE and MAE for the proposed forecasts: exact conditional median $X_1^{(\text{Mn})}(1)$, approximate median $X_1^{(\text{Map})}(1)$ and conditional mode $X_1^{(\text{Md})}(1)$ considering known parameters. Notice that in all cases the RMSE and the MAE of the exact conditional median and its approximation are equal, and they are slightly smaller than the RMSE and MAE of conditional mode, respectively.

Therefore, for known parameters, the approximate median $X_1^{(\text{Map})}(1)$ provides the same results of the conditional median $X_1^{(\text{Mn})}(1)$, and they are slightly better that the conditional mode $X_1^{(\text{Md})}(1)$, in terms of RMSE and MAE.

In the second scenario we consider unknown parameters. We simulate Monte Carlo samples of size $N + T + 1$ and use the last $T + 1$ elements of each sample; this is performed to guarantee that the sample of size $T + 1$ starts approximately in the marginal stationary distribution. After that, we estimate the parameters α and λ for each sample using YW, CLS and CML estimation methods. For each sample, we verify if the parameter estimates are in parametric space; if it does not happen, then, we discard the sample and substitute it by another Monte Carlo sample; for each valid sample we find the forecasts $X_T^{(\text{Mn})}(1)$, $X_T^{(\text{Map})}(1)$ and $X_T^{(\text{Md})}(1)$.

Repeat until $r = 10000$ and for $N = 1500$

1. $y_1^{(r)} = 0$

2. $y_t^{(r)} = \text{Po}(\alpha y_{t-1}^{(r)} + \lambda), \quad t = 2, \ldots, N+T+1$

   $\dashrightarrow \quad y_1^{(r)}, \ldots, y_N^{(r)}, y_{N+1}^{(r)}, \ldots, y_{N+T+1}^{(r)}$

   $\dashrightarrow \quad x_i^{(r)} = y_{N+i}^{(r)} \quad \text{for} \quad i = 1, \ldots, T+1$

   $\dashrightarrow \quad x_1^{(r)}, \ldots, x_T^{(r)}, x_{T+1}^{(r)}$

3. $x_1^{(r)}, x_2^{(r)}, \ldots x_T^{(r)} \quad \dashrightarrow \quad (\widehat{\alpha}_{\text{YW}}, \widehat{\lambda}_{\text{YW}}), (\widehat{\alpha}_{\text{CLS}}, \widehat{\lambda}_{\text{CLS}}), (\widehat{\alpha}_{\text{CML}}, \widehat{\lambda}_{\text{CML}})$

4. If $(\widehat{\alpha}_{\text{YW}}, \widehat{\lambda}_{\text{YW}}), (\widehat{\alpha}_{\text{CLS}}, \widehat{\lambda}_{\text{CLS}}) \in (0,1) \times (0, \infty)$

   $\dashrightarrow \quad X_T^{(\text{Mn})}(1)^{(r)},\ X_T^{(\text{Map})}(1)^{(r)}$ and $X_T^{(\text{Md})}(1)^{(r)}; \quad r = r+1$ and return to step 1

   Else

   $\dashrightarrow$ return to step 1 without calculating forecasts and without updating r.

Tables 4.2 and 4.3 show the results for unknown parameters. For $\alpha = 0.1, 0.5$, $T = 25$, 50 and for all λ considered the conditional median and its approximation were competitive and slightly better than the conditional mode, while for $\alpha = 0.9$ the three forecasts were competitive and YW estimators were a little worse than the others, CML estimators being slightly better than CLS, in terms of RMSE and MAE. For $T = 100$ the three estimation methods and the three forecasts were competitive, respectively, in terms of RMSE and MAE.

In summary, for unknown parameters, the conditional median $X_T^{(\text{Mn})}(1)$ and its approximation $X_T^{(\text{Map})}(1)$ provided almost the same results, and they were slightly better that the conditional mode $X_T^{(\text{Md})}(1)$ for small and moderated sample sizes, in terms of RMSE and MAE. YW estimators were a little worse than the others for small and moderated sample sizes, in terms of RSME and MAE. For large T the estimation methods and the forecasts considered provided almost the same results, in terms of RMSE and MAE.

Therefore we suggest to use the approximated conditional median and CLS estimators because, the approximated conditional median has an easy analytical expression and

CLS and CML estimators produce almost the same results but CLS estimators have the advantage of being found more easily than CML estimators, they have explicit formulas while CML estimators are found by numerical maximization.

	Error	Forecast	α = 0.1	0.5	0.9
$\lambda = 0.5$	RMSE	$X_1^{(\text{Mn})}(1)$	0.899	1.064	2.240
		$X_1^{(\text{Map})}(1)$	0.899	1.064	2.240
		$X_1^{(\text{Md})}(1)$	0.926	1.064	2.264
	MAE	$X_1^{(\text{Mn})}(1)$	0.550	0.704	1.575
		$X_1^{(\text{Map})}(1)$	0.550	0.704	1.575
		$X_1^{(\text{Md})}(1)$	0.553	0.704	1.581
$\lambda = 1$	RMSE	$X_1^{(\text{Mn})}(1)$	1.047	1.468	3.150
		$X_1^{(\text{Map})}(1)$	1.047	1.468	3.150
		$X_1^{(\text{Md})}(1)$	1.047	1.468	3.173
	MAE	$X_1^{(\text{Mn})}(1)$	0.763	1.080	2.350
		$X_1^{(\text{Map})}(1)$	0.763	1.080	2.350
		$X_1^{(\text{Md})}(1)$	0.763	1.080	2.365
$\lambda = 3$	RMSE	$X_1^{(\text{Mn})}(1)$	1.865	2.476	5.510
		$X_1^{(\text{Map})}(1)$	1.865	2.476	5.510
		$X_1^{(\text{Md})}(1)$	1.870	2.476	5.526
	MAE	$X_1^{(\text{Mn})}(1)$	1.430	1.914	4.297
		$X_1^{(\text{Map})}(1)$	1.430	1.914	4.297
		$X_1^{(\text{Md})}(1)$	1.432	1.914	4.305
$\lambda = 5$	RMSE	$X_1^{(\text{Mn})}(1)$	2.385	3.167	7.123
		$X_1^{(\text{Map})}(1)$	2.385	3.167	7.123
		$X_1^{(\text{Md})}(1)$	2.415	3.167	7.145
	MAE	$X_1^{(\text{Mn})}(1)$	1.860	2.489	5.636
		$X_1^{(\text{Map})}(1)$	1.860	2.489	5.636
		$X_1^{(\text{Md})}(1)$	1.867	2.489	5.651

Table 4.1: RMSE and MAE of $X_1^{(\text{Mn})}(1)$, $X_1^{(\text{Map})}(1)$ and $X_1^{(\text{Md})}(1)$, for different values of known parameters α and λ.

	Error	Forecast	Estimator	T = 25			T = 50			T = 100		
				α			α			α		
				0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
$\lambda = 0.5$	RMSE	$X_T^{(\mathrm{Mn})}(1)$	YW	0.89	1.08	2.44	0.90	1.07	2.36	0.92	1.08	2.31
		$X_T^{(\mathrm{Mn})}(1)$	CML	0.89	1.09	2.31	0.90	1.07	2.29	0.92	1.07	2.28
		$X_T^{(\mathrm{Mn})}(1)$	CLS	0.89	1.09	2.33	0.90	1.07	2.30	0.92	1.08	2.29
		$X_T^{(\mathrm{Map})}(1)$	YW	0.89	1.08	2.44	0.90	1.07	2.36	0.91	1.08	2.31
		$X_T^{(\mathrm{Map})}(1)$	CML	0.89	1.08	2.31	0.90	1.06	2.29	0.91	1.07	2.28
		$X_T^{(\mathrm{Map})}(1)$	CLS	0.89	1.08	2.33	0.90	1.07	2.30	0.91	1.08	2.29
		$X_T^{(\mathrm{Md})}(1)$	YW	0.92	1.17	2.48	0.92	1.17	2.41	0.94	1.16	2.36
		$X_T^{(\mathrm{Md})}(1)$	CML	0.92	1.16	2.34	0.92	1.16	2.33	0.93	1.16	2.31
		$X_T^{(\mathrm{Md})}(1)$	CLS	0.92	1.16	2.36	0.92	1.16	2.35	0.94	1.16	2.33
	MAE	$X_T^{(\mathrm{Mn})}(1)$	YW	0.57	0.74	1.75	0.57	0.73	1.68	0.56	0.72	1.62
		$X_T^{(\mathrm{Mn})}(1)$	CML	0.57	0.74	1.64	0.57	0.73	1.60	0.56	0.72	1.59
		$X_T^{(\mathrm{Mn})}(1)$	CLS	0.57	0.74	1.66	0.57	0.73	1.63	0.56	0.73	1.60
		$X_T^{(\mathrm{Map})}(1)$	YW	0.58	0.74	1.75	0.57	0.73	1.68	0.57	0.73	1.62
		$X_T^{(\mathrm{Map})}(1)$	CML	0.58	0.74	1.64	0.57	0.72	1.60	0.57	0.72	1.59
		$X_T^{(\mathrm{Map})}(1)$	CLS	0.58	0.74	1.66	0.57	0.73	1.63	0.57	0.73	1.60
		$X_T^{(\mathrm{Md})}(1)$	YW	0.55	0.78	1.76	0.56	0.78	1.68	0.55	0.77	1.62
		$X_T^{(\mathrm{Md})}(1)$	CML	0.55	0.77	1.65	0.55	0.77	1.62	0.55	0.77	1.59
		$X_T^{(\mathrm{Md})}(1)$	CLS	0.55	0.77	1.67	0.56	0.78	1.64	0.55	0.77	1.61
$\lambda = 1$	RMSE	$X_T^{(\mathrm{Mn})}(1)$	YW	1.10	1.52	3.51	1.08	1.48	3.31	1.06	1.46	3.30
		$X_T^{(\mathrm{Mn})}(1)$	CML	1.11	1.52	3.34	1.09	1.48	3.24	1.06	1.46	3.26
		$X_T^{(\mathrm{Mn})}(1)$	CLS	1.11	1.53	3.35	1.09	1.48	3.25	1.06	1.46	3.26
		$X_T^{(\mathrm{Map})}(1)$	YW	1.10	1.52	3.51	1.08	1.47	3.31	1.05	1.46	3.30
		$X_T^{(\mathrm{Map})}(1)$	CML	1.11	1.52	3.34	1.08	1.47	3.24	1.05	1.46	3.26
		$X_T^{(\mathrm{Map})}(1)$	CLS	1.11	1.52	3.35	1.08	1.48	3.25	1.06	1.46	3.26
		$X_T^{(\mathrm{Md})}(1)$	YW	1.23	1.59	3.54	1.22	1.56	3.35	1.17	1.54	3.33
		$X_T^{(\mathrm{Md})}(1)$	CML	1.24	1.58	3.36	1.23	1.55	3.26	1.17	1.54	3.29
		$X_T^{(\mathrm{Md})}(1)$	CLS	1.24	1.59	3.37	1.23	1.55	3.27	1.17	1.54	3.30
	MAE	$X_T^{(\mathrm{Mn})}(1)$	YW	0.80	1.13	2.64	0.79	1.09	2.49	0.76	1.08	2.46
		$X_T^{(\mathrm{Mn})}(1)$	CML	0.81	1.13	2.50	0.79	1.09	2.43	0.76	1.08	2.44
		$X_T^{(\mathrm{Mn})}(1)$	CLS	0.81	1.14	2.51	0.79	1.09	2.44	0.76	1.07	2.44
		$X_T^{(\mathrm{Map})}(1)$	YW	0.80	1.13	2.64	0.79	1.09	2.49	0.76	1.07	2.46
		$X_T^{(\mathrm{Map})}(1)$	CML	0.81	1.13	2.50	0.79	1.09	2.43	0.76	1.08	2.44
		$X_T^{(\mathrm{Map})}(1)$	CLS	0.81	1.13	2.51	0.79	1.09	2.44	0.76	1.07	2.44
		$X_T^{(\mathrm{Md})}(1)$	YW	0.88	1.16	2.65	0.87	1.12	2.50	0.83	1.12	2.47
		$X_T^{(\mathrm{Md})}(1)$	CML	0.89	1.16	2.50	0.88	1.13	2.44	0.83	1.12	2.45
		$X_T^{(\mathrm{Md})}(1)$	CLS	0.88	1.16	2.51	0.88	1.12	2.45	0.83	1.12	2.45

Table 4.2: RMSE and MAE of $X_T^{(\mathrm{Mn})}(1)$, $X_T^{(\mathrm{Map})}(1)$ and $X_T^{(\mathrm{Md})}(1)$ for different values of α and $\lambda = 0.5, 1$, using three estimation methods, YW, CML and CLS, for sample sizes $T = 25, 50$ and 100.

	Error	Forecast	Estimator	$T = 25$			$T = 50$			$T = 100$		
				α			α			α		
				0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
$\lambda = 3$	RMSE	$X_T^{(\mathrm{Mn})}(1)$	YW	1.89	2.62	6.07	1.89	2.57	5.74	1.89	2.48	5.56
		$X_T^{(\mathrm{Mn})}(1)$	CML	1.90	2.62	5.78	1.89	2.57	5.60	1.90	2.49	5.51
		$X_T^{(\mathrm{Mn})}(1)$	CLS	1.90	2.62	5.78	1.89	2.57	5.60	1.90	2.49	5.52
		$X_T^{(\mathrm{Map})}(1)$	YW	1.89	2.62	6.07	1.89	2.57	5.74	1.89	2.48	5.56
		$X_T^{(\mathrm{Map})}(1)$	CML	1.90	2.62	5.78	1.89	2.57	5.60	1.90	2.49	5.51
		$X_T^{(\mathrm{Map})}(1)$	CLS	1.90	2.62	5.78	1.89	2.57	5.60	1.90	2.48	5.52
		$X_T^{(\mathrm{Md})}(1)$	YW	1.94	2.65	6.07	1.95	2.61	5.76	1.93	2.52	5.57
		$X_T^{(\mathrm{Md})}(1)$	CML	1.96	2.65	5.79	1.95	2.60	5.62	1.93	2.52	5.53
		$X_T^{(\mathrm{Md})}(1)$	CLS	1.96	2.65	5.79	1.95	2.60	5.62	1.93	2.53	5.53
	MAE	$X_T^{(\mathrm{Mn})}(1)$	YW	1.47	2.04	4.75	1.46	1.99	4.50	1.45	1.93	4.35
		$X_T^{(\mathrm{Mn})}(1)$	CML	1.48	2.04	4.53	1.46	1.99	4.38	1.45	1.94	4.31
		$X_T^{(\mathrm{Mn})}(1)$	CLS	1.48	2.04	4.53	1.46	1.99	4.38	1.45	1.93	4.32
		$X_T^{(\mathrm{Map})}(1)$	YW	1.47	2.04	4.75	1.46	2.00	4.50	1.45	1.93	4.35
		$X_T^{(\mathrm{Map})}(1)$	CML	1.48	2.04	4.53	1.46	1.99	4.38	1.45	1.94	4.31
		$X_T^{(\mathrm{Map})}(1)$	CLS	1.48	2.04	4.53	1.46	1.99	4.38	1.45	1.93	4.32
		$X_T^{(\mathrm{Md})}(1)$	YW	1.49	2.04	4.74	1.48	2.01	4.51	1.46	1.95	4.35
		$X_T^{(\mathrm{Md})}(1)$	CML	1.50	2.05	4.53	1.48	2.01	4.40	1.46	1.95	4.32
		$X_T^{(\mathrm{Md})}(1)$	CLS	1.50	2.05	4.53	1.48	2.00	4.39	1.46	1.95	4.32
$\lambda = 5$	RMSE	$X_T^{(\mathrm{Mn})}(1)$	YW	2.44	3.29	7.90	2.40	3.23	7.33	2.40	3.17	7.19
		$X_T^{(\mathrm{Mn})}(1)$	CML	2.45	3.29	7.51	2.41	3.23	7.19	2.41	3.17	7.14
		$X_T^{(\mathrm{Mn})}(1)$	CLS	2.45	3.30	7.51	2.41	3.23	7.19	2.41	3.17	7.14
		$X_T^{(\mathrm{Map})}(1)$	YW	2.45	3.29	7.90	2.40	3.23	7.33	2.40	3.17	7.19
		$X_T^{(\mathrm{Map})}(1)$	CML	2.45	3.29	7.51	2.41	3.23	7.19	2.41	3.17	7.14
		$X_T^{(\mathrm{Map})}(1)$	CLS	2.45	3.30	7.52	2.41	3.23	7.19	2.41	3.17	7.14
		$X_T^{(\mathrm{Md})}(1)$	YW	2.48	3.33	7.92	2.45	3.27	7.34	2.44	3.21	7.20
		$X_T^{(\mathrm{Md})}(1)$	CML	2.49	3.33	7.52	2.45	3.27	7.20	2.44	3.20	7.15
		$X_T^{(\mathrm{Md})}(1)$	CLS	2.49	3.33	7.53	2.45	3.27	7.20	2.44	3.20	7.15
	MAE	$X_T^{(\mathrm{Mn})}(1)$	YW	1.90	2.60	6.19	1.88	2.54	5.80	1.88	2.51	5.66
		$X_T^{(\mathrm{Mn})}(1)$	CML	1.91	2.59	5.89	1.89	2.55	5.68	1.88	2.51	5.63
		$X_T^{(\mathrm{Mn})}(1)$	CLS	1.91	2.60	5.89	1.89	2.55	5.68	1.88	2.51	5.63
		$X_T^{(\mathrm{Map})}(1)$	YW	1.90	2.60	6.19	1.88	2.54	5.80	1.88	2.51	5.66
		$X_T^{(\mathrm{Map})}(1)$	CML	1.91	2.59	5.89	1.89	2.55	5.68	1.88	2.51	5.63
		$X_T^{(\mathrm{Map})}(1)$	CLS	1.91	2.60	5.89	1.89	2.55	5.68	1.88	2.51	5.63
		$X_T^{(\mathrm{Md})}(1)$	YW	1.91	2.61	6.19	1.90	2.56	5.80	1.89	2.53	5.65
		$X_T^{(\mathrm{Md})}(1)$	CML	1.92	2.61	5.90	1.90	2.56	5.69	1.89	2.52	5.63
		$X_T^{(\mathrm{Md})}(1)$	CLS	1.92	2.62	5.91	1.90	2.56	5.69	1.89	2.52	5.63

Table 4.3: RMSE and MAE of $X_T^{(\mathrm{Mn})}(1)$, $X_T^{(\mathrm{Map})}(1)$ and $X_T^{(\mathrm{Md})}(1)$ for different values of α and $\lambda = 3, 5$, using three estimation methods, YW, CML and CLS, for sample sizes $T =$ 25, 50 and 100.

4.8 Monte Carlo results for forecasting two-steps ahead

In this section we present a Monte Carlo study that compares the behavior of the following forecasts: the conditional median $X_T^{(\mathrm{Mn})}(2)$, the conditional mode $X_T^{(\mathrm{Md})}(2)$, the forecast $X_T^{(\mathrm{Map})}(2) = \left\lceil \hat{\alpha} X_T^{(\mathrm{Map})}(1) + \hat{\lambda} - \frac{2}{3} \right\rceil$ proposed in equation (4.37) with $h = 2$ as well as the integer part of conditional mean two-steps ahead, $X_T^{(\mathrm{Ei})}(2)$, proposed in equation (4.36). We use CLS and CML estimation methods.

We simulate $R = 10000$ valid Monte Carlo samples $X_1, X_2, \ldots, X_T, X_{T+1}, X_{T+2}$ using the procedure explained in Section 4.7 and, then, we estimate the parameters α and λ; for each sample we find the forecasts $X_T^{(\mathrm{Mn})}(2)$, $X_T^{(\mathrm{Md})}(2)$, $X_T^{(\mathrm{Map})}(2)$ and $X_T^{(\mathrm{Ei})}(2)$ using the estimates of α and λ provided by each estimation method. Finally we find the simulated MAE and the simulated RMSE of the four forecasts. Table 4.4 and Table 4.5 show these results.

In general CLS and CML estimators provided almost the same results, CML estimators being slightly better than CLS estimators, and the conditional mode $X_T^{(\mathrm{Md})}(2)$ was worse than the other forecasts, in terms of RMSE and MAE.

For $\lambda = 1$, $\alpha = 0.1, 0.5, 0.9$, $T = 25$ and $T = 50$ the recursive expression $X_T^{(\mathrm{Map})}(2)$ and the conditional median $X_T^{(\mathrm{Mn})}(2)$ were competitive and they were slightly better than the integer part of the conditional mean $X_T^{(\mathrm{Ei})}(2)$, while for $T = 100$ the forecasts $X_T^{(\mathrm{Mn})}(2)$, $X_T^{(\mathrm{Map})}(2)$ and $X_T^{(\mathrm{Ei})}(2)$ provided almost the same results, in terms of RMSE and MAE. For $\lambda = 3$, $\alpha = 0.1, 0.5$, $T = 25$ and $T = 50$ the forecasts $X_T^{(\mathrm{Mn})}(2)$ and $X_T^{(\mathrm{Map})}(2)$ were competitive and they were slightly better than the forecast $X_T^{(\mathrm{Ei})}(2)$, while for $\alpha = 0.9$ the forecasts $X_T^{(\mathrm{Map})}(2)$ and $X_T^{(\mathrm{Ei})}(2)$ were competitive and they were slightly better than $X_T^{(\mathrm{Mn})}(2)$, in terms of RMSE and MAE. For $T = 100$ the forecasts $X_T^{(\mathrm{Mn})}(2)$, $X_T^{(\mathrm{Map})}(2)$ and $X_T^{(\mathrm{Ei})}(2)$ provided almost the same results, in terms of RMSE and MAE. For $\lambda = 5$, $\alpha = 0.1$, $T = 25$ and $T = 50$ the forecasts $X_T^{(\mathrm{Mn})}(2)$ and $X_T^{(\mathrm{Map})}(2)$ were competitive and they were slightly better than the forecast $X_T^{(\mathrm{Ei})}(2)$, for $\alpha = 0.5$, $T = 25$ and $T = 50$ the forecasts $X_T^{(\mathrm{Mn})}(2)$, $X_T^{(\mathrm{Map})}(2)$ and $X_T^{(\mathrm{Md})}(2)$ were competitive, while for $\alpha = 0.9$, $T = 25$, 50 and $T = 100$ the forecasts $X_T^{(\mathrm{Map})}(2)$ and $X_T^{(\mathrm{Ei})}(2)$ were competitive and they were slightly better than the forecast $X_T^{(\mathrm{Mn})}(2)$.

In summary, the conditional mode $X_T^{(\mathrm{Md})}(2)$ was the worst forecast, the forecasts $X_T^{(\mathrm{Map})}(2)$

and $X_T^{(\mathrm{Mn})}(2)$ provided almost the same results and for large T the forecasts $X_T^{(\mathrm{Map})}(2)$, $X_T^{(\mathrm{Mn})}(2)$ and $X_T^{(\mathrm{Ei})}(2)$ were competitive, in terms of RMSE and MAE. CLS and CML estimators provided almost the same results, CML estimators being slightly better than CLS estimators, in terms of RMSE and MAE.

Therefore, we suggest to use $X_T^{(\mathrm{Map})}(2)$ and CLS estimators, because in general $X_T^{(\mathrm{Map})}(2)$ was slightly better than the others forecasts and CLS estimators have an easy closed form, while the CML estimators are found by using complicated maximization methods and they produce almost the same results than the CLS estimators.

4.9 Monte Carlo results for forecasting $h-$steps ahead

In this section we show a Monte Carlo simulation that studies the behavior of

$$X_T^{(\mathrm{Map})}(h) = \begin{cases} \left\lceil \widehat{\alpha}_{\mathrm{CLS}}\, X_T + \widehat{\lambda}_{\mathrm{CLS}} - \frac{2}{3} \right\rceil & \text{if} \quad h = 1 \\ \left\lceil \widehat{\alpha}_{\mathrm{CLS}}\, X_T^{(\mathrm{Map})}(h-1) + \widehat{\lambda}_{\mathrm{CLS}} - \frac{2}{3} \right\rceil & \text{if} \quad h \geqslant 2, \end{cases}$$

and

$$X_T^{(\mathrm{Ei})}(h) = \left\lfloor \widehat{\alpha}_{\mathrm{CLS}}^{h} X_T + \left(\frac{1-\widehat{\alpha}_{\mathrm{CLS}}^{h}}{1-\widehat{\alpha}_{\mathrm{CLS}}} \right) \widehat{\lambda}_{\mathrm{CLS}} \right\rfloor$$

as forecasts of X_{T+h}, where $\widehat{\alpha}_{\mathrm{CLS}}$ and $\widehat{\lambda}_{\mathrm{CLS}}$ represent the CLS estimators of α and λ respectively. Notice that, for $h = 1$, $X_T^{(\mathrm{Map})}(1)$ is the forecast of X_{T+1} proposed in Equation 4.12.

For the simulation study, we generate 10000 valid Monte Carlo samples as explained in Section 4.7, considering different values of α and λ, namely, α = 0.1, 0.5, 0.9 and λ = 0.5, 1, 3, 5 and we considered the forecasts h = 1, 2, . . . , 10 steps ahead. Finally, we took the sample sizes as T = 25, 50 and 100.

Tables 4.6, 4.7, 4.8, 4.9, 4.10 and 4.11 show the RMSE and MAE of $X_T^{(\mathrm{Map})}(h)$ and $X_T^{(\mathrm{Ei})}(h)$ for $h = 1, 2, \ldots, 10$ and for the different values of α and λ considered. Note that for $\alpha = 0.1$ and for each λ the RMSE of the forecasts $X_T^{(\mathrm{Map})}(h)$ and $X_T^{(\mathrm{Ei})}(h)$ for $h = 1, 2, \ldots, 10$ are very close for all sample sizes considered; the same happens for the MAE. For $\alpha = 0.5$ and for each λ the RMSE of the forecasts $X_T^{(\mathrm{Map})}(h)$ and $X_T^{(\mathrm{Ei})}(h)$ for

	Error	Estimator	Forecast	T = 25			T = 50			T = 100		
				α			α			α		
				0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
$\lambda = 0.5$	RMSE	CLS	$X_T^{(\mathrm{Mn})}(2)$	0.923	1.264	3.247	0.920	1.220	3.093	0.933	1.203	3.093
			$X_T^{(\mathrm{Md})}(2)$	0.937	1.386	3.354	0.929	1.405	3.246	0.938	1.432	3.252
			$X_T^{(\mathrm{Map})}(2)$	0.915	1.270	3.242	0.915	1.237	3.105	0.925	1.239	3.105
			$X_T^{(\mathrm{Ei})}(2)$	0.935	1.336	3.266	0.929	1.326	3.109	0.938	1.326	3.109
		CML	$X_T^{(\mathrm{Mn})}(2)$	0.923	1.268	3.227	0.919	1.221	3.071	0.933	1.207	3.076
			$X_T^{(\mathrm{Md})}(2)$	0.936	1.385	3.348	0.929	1.407	3.232	0.938	1.434	3.250
			$X_T^{(\mathrm{Map})}(2)$	0.915	1.269	3.225	0.915	1.242	3.070	0.926	1.240	3.094
			$X_T^{(\mathrm{Ei})}(2)$	0.935	1.334	3.249	0.929	1.325	3.088	0.938	1.324	3.093
	MAE	CLS	$X_T^{(\mathrm{Mn})}(2)$	0.588	0.881	2.311	0.565	0.858	2.206	0.571	0.849	2.180
			$X_T^{(\mathrm{Md})}(2)$	0.558	0.915	2.372	0.549	0.936	2.284	0.564	0.953	2.263
			$X_T^{(\mathrm{Map})}(2)$	0.593	0.879	2.301	0.576	0.865	2.218	0.579	0.857	2.197
			$X_T^{(\mathrm{Ei})}(2)$	0.558	0.897	2.313	0.550	0.892	2.206	0.564	0.894	2.181
		CML	$X_T^{(\mathrm{Mn})}(2)$	0.589	0.883	2.288	0.565	0.859	2.186	0.571	0.848	2.171
			$X_T^{(\mathrm{Md})}(2)$	0.558	0.912	2.366	0.549	0.936	2.279	0.564	0.955	2.271
			$X_T^{(\mathrm{Map})}(2)$	0.594	0.878	2.280	0.575	0.865	2.189	0.579	0.857	2.183
			$X_T^{(\mathrm{Ei})}(2)$	0.560	0.894	2.297	0.549	0.893	2.189	0.564	0.892	2.172
$\lambda = 1$	RMSE	CLS	$X_T^{(\mathrm{Mn})}(2)$	1.086	1.722	4.591	1.058	1.665	4.428	1.060	1.658	4.385
			$X_T^{(\mathrm{Md})}(2)$	1.259	1.824	4.658	1.210	1.775	4.505	1.177	1.798	4.484
			$X_T^{(\mathrm{Map})}(2)$	1.103	1.736	4.594	1.069	1.677	4.436	1.062	1.673	4.387
			$X_T^{(\mathrm{Ei})}(2)$	1.227	1.771	4.602	1.184	1.715	4.434	1.153	1.718	4.395
		CML	$X_T^{(\mathrm{Mn})}(2)$	1.090	1.720	4.586	1.058	1.664	4.419	1.061	1.660	4.373
			$X_T^{(\mathrm{Md})}(2)$	1.259	1.824	4.648	1.211	1.775	4.506	1.179	1.796	4.476
			$X_T^{(\mathrm{Map})}(2)$	1.108	1.739	4.590	1.067	1.676	4.434	1.062	1.673	4.389
			$X_T^{(\mathrm{Ei})}(2)$	1.229	1.770	4.589	1.185	1.709	4.422	1.154	1.718	4.387
	MAE	CLS	$X_T^{(\mathrm{Mn})}(2)$	0.790	1.278	3.467	0.764	1.247	3.352	0.761	1.231	3.307
			$X_T^{(\mathrm{Md})}(2)$	0.898	1.322	3.514	0.864	1.288	3.380	0.832	1.287	3.346
			$X_T^{(\mathrm{Map})}(2)$	0.801	1.291	3.471	0.771	1.250	3.359	0.762	1.238	3.317
			$X_T^{(\mathrm{Ei})}(2)$	0.875	1.294	3.469	0.846	1.258	3.347	0.815	1.246	3.306
		CML	$X_T^{(\mathrm{Mn})}(2)$	0.793	1.278	3.465	0.764	1.244	3.341	0.761	1.235	3.298
			$X_T^{(\mathrm{Md})}(2)$	0.898	1.323	3.506	0.864	1.286	3.375	0.833	1.286	3.347
			$X_T^{(\mathrm{Map})}(2)$	0.803	1.294	3.468	0.770	1.247	3.360	0.762	1.235	3.320
			$X_T^{(\mathrm{Ei})}(2)$	0.877	1.294	3.463	0.847	1.251	3.338	0.817	1.248	3.302

Table 4.4: RMSE and MAE of $X_T^{(\mathrm{Mn})}(2)$, $X_T^{(\mathrm{Md})}(2)$, $X_T^{(\mathrm{Map})}(2)$ and $X_T^{(\mathrm{Ei})}(2)$ for different values of α and $\lambda = 0.5$, 1, using CLS and CML estimation methods, for sample sizes $T = 25$, 50 and 100.

	Error	Estimator	Forecast	$T=25$			$T=50$			$T=100$		
				α			α			α		
				0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
$\lambda=3$	RMSE	CLS	$X_T^{(\mathrm{Mn})}(2)$	1.915	2.911	7.916	1.885	2.872	7.713	1.853	2.820	7.542
			$X_T^{(\mathrm{Md})}(2)$	1.976	2.983	8.098	1.944	2.948	7.902	1.877	2.904	7.727
			$X_T^{(\mathrm{Map})}(2)$	1.915	2.923	7.940	1.887	2.877	7.751	1.852	2.829	7.525
			$X_T^{(\mathrm{Ei})}(2)$	1.966	2.950	7.945	1.935	2.901	7.741	1.872	2.851	7.519
		CML	$X_T^{(\mathrm{Mn})}(2)$	1.914	2.907	7.908	1.886	2.872	7.705	1.853	2.818	7.541
			$X_T^{(\mathrm{Md})}(2)$	1.978	2.977	8.101	1.945	2.947	7.901	1.877	2.902	7.722
			$X_T^{(\mathrm{Map})}(2)$	1.916	2.914	7.939	1.887	2.878	7.741	1.852	2.829	7.508
			$X_T^{(\mathrm{Ei})}(2)$	1.967	2.944	7.944	1.936	2.899	7.735	1.872	2.850	7.508
	MAE	CLS	$X_T^{(\mathrm{Mn})}(2)$	1.467	2.269	6.191	1.446	2.242	6.006	1.421	2.197	5.869
			$X_T^{(\mathrm{Md})}(2)$	1.500	2.306	6.295	1.474	2.267	6.095	1.430	2.236	5.967
			$X_T^{(\mathrm{Map})}(2)$	1.468	2.279	6.205	1.448	2.241	6.029	1.420	2.205	5.874
			$X_T^{(\mathrm{Ei})}(2)$	1.494	2.285	6.204	1.468	2.246	6.015	1.429	2.205	5.864
		CML	$X_T^{(\mathrm{Mn})}(2)$	1.467	2.267	6.188	1.447	2.243	5.993	1.420	2.195	5.870
			$X_T^{(\mathrm{Md})}(2)$	1.502	2.302	6.294	1.474	2.266	6.093	1.431	2.232	5.967
			$X_T^{(\mathrm{Map})}(2)$	1.469	2.276	6.205	1.447	2.241	6.016	1.420	2.204	5.864
			$X_T^{(\mathrm{Ei})}(2)$	1.496	2.285	6.203	1.470	2.246	6.006	1.428	2.201	5.859
$\lambda=5$	RMSE	CLS	$X_T^{(\mathrm{Mn})}(2)$	2.466	3.747	11.120	2.468	3.534	10.951	2.415	3.556	11.040
			$X_T^{(\mathrm{Md})}(2)$	2.516	3.790	11.413	2.506	3.593	11.165	2.429	3.629	11.317
			$X_T^{(\mathrm{Map})}(2)$	2.468	3.751	10.339	2.469	3.545	9.936	2.413	3.557	9.812
			$X_T^{(\mathrm{Ei})}(2)$	2.509	3.767	10.332	2.500	3.561	9.929	2.427	3.590	9.815
		CML	$X_T^{(\mathrm{Mn})}(2)$	2.466	3.745	11.119	2.470	3.535	10.942	2.415	3.552	11.041
			$X_T^{(\mathrm{Md})}(2)$	2.517	3.793	11.427	2.507	3.592	11.177	2.429	3.627	11.330
			$X_T^{(\mathrm{Map})}(2)$	2.470	3.747	10.331	2.469	3.547	9.931	2.414	3.556	9.800
			$X_T^{(\mathrm{Ei})}(2)$	2.507	3.767	10.332	2.501	3.559	9.933	2.427	3.587	9.803
	MAE	CLS	$X_T^{(\mathrm{Mn})}(2)$	1.926	2.949	8.524	1.927	2.763	8.347	1.883	2.800	8.405
			$X_T^{(\mathrm{Md})}(2)$	1.948	2.966	8.767	1.930	2.785	8.554	1.883	2.835	8.660
			$X_T^{(\mathrm{Map})}(2)$	1.929	2.951	8.134	1.926	2.769	7.830	1.881	2.799	7.752
			$X_T^{(\mathrm{Ei})}(2)$	1.946	2.956	8.122	1.927	2.771	7.814	1.882	2.814	7.746
		CML	$X_T^{(\mathrm{Mn})}(2)$	1.925	2.947	8.517	1.928	2.766	8.339	1.884	2.798	8.403
			$X_T^{(\mathrm{Md})}(2)$	1.949	2.969	8.767	1.931	2.783	8.556	1.883	2.835	8.666
			$X_T^{(\mathrm{Map})}(2)$	1.930	2.948	8.125	1.926	2.770	7.825	1.883	2.801	7.741
			$X_T^{(\mathrm{Ei})}(2)$	1.945	2.957	8.119	1.927	2.768	7.817	1.882	2.814	7.740

Table 4.5: RMSE and MAE of $X_T^{(\mathrm{Mn})}(2)$, $X_T^{(\mathrm{Md})}(2)$, $X_T^{(\mathrm{Map})}(2)$ and $X_T^{(\mathrm{Ei})}(2)$ for different values of α and $\lambda = 3, 5$, using CLS and CML estimation methods, for sample sizes $T = 25, 50$ and 100.

$h = 1, 2, \ldots, 10$ increases a little, as the number of steps ahead, h, increases from 1 to 10; a similar effect happens with the MAE. For $\alpha = 0.9$ and for each λ the RMSE of the forecasts $X_T^{(\text{Map})}(h)$ and $X_T^{(\text{Ei})}(h)$ for $h = 1, 2, \ldots, 10$ increases considerably as h increases from 1 to 10; the same is true for the MAE.
Furthermore, in terms of RMSE, for $\alpha = 0.1$ and $\alpha = 0.5$ and for all values of λ, either $X_T^{(\text{Map})}(h)$ is better than $X_T^{(\text{Ei})}(h)$ or they are competitive, the same happens for $\alpha = 0.9$ and the sample sizes $T = 25$ and $T = 50$. However for $T = 100$ and for large values of h, $X_T^{(\text{Ei})}(h)$ is a little better than $X_T^{(\text{Map})}(h)$. On the other hand, in terms of MAE, for $\alpha = 0.1$ and for λ = 1, 2, 3, the forecasts $X_T^{(\text{Ei})}(h)$ and $X_T^{(\text{Map})}(h)$ are competitive, $X_T^{(\text{Map})}(h)$ being a little better; for $\lambda = 0.5$, $X_T^{(\text{Ei})}(h)$ is a little better than $X_T^{(\text{Map})}(h)$. For $\alpha = 0.5$ and sample sizes $T = 50$ and $T = 100$, the forecasts $X_T^{(\text{Ei})}(h)$ and $X_T^{(\text{Map})}(h)$ are competitive, $X_T^{(\text{Map})}(h)$ being a little better, for λ = 1, 3, $T = 25$ and for large values of h, $X_T^{(\text{Ei})}(h)$ is a little better than $X_T^{(\text{Map})}(h)$. Finally, for $\alpha = 0.9$, the forecasts $X_T^{(\text{Ei})}(h)$ and $X_T^{(\text{Map})}(h)$ are competitive, $X_T^{(\text{Map})}(h)$ being a little better, except for the cases λ =0.5 , 1, $T = 100$ and for large values of h; in these cases $X_T^{(\text{Ei})}(h)$ is a little better than $X_T^{(\text{Map})}(h)$.

Therefore the predictors $X_T^{(\text{Map})}(h)$ and $X_T^{(\text{Ei})}(h)$ have similar behavior and we are not able to conclude that none is better than the other. Finally, we find that for small α the RMSE and the MAE of the forecast one-step ahead are not much smaller than of the forecasts $h = 2, \ldots, 10$ steps ahead, while for large α the RMSE and the MAE increase considerably from one-step ahead to two-steps ahead and then increases more slowly, for both predictors. For $X_T^{(\text{Map})}(h)$, this can be explained from the fact that the forecasts one-step ahead use the conditional distribution, but the forecasts $h \geqslant 2$ steps ahead only use a recursive expression without using the conditional distribution h steps ahead.

4.10 Applications

In this section two data sets are used to illustrate and compare $X_T^{(\text{Mn})}(h)$, $X_T^{(\text{Md})}(h)$ and $X_T^{(\text{Map})}(h)$ for $h = 1$ and $h = 2$ in the INARCH(1) process.

The first application considered is the monthly number of Polio cases in the United States between January 1970 until December 1983, a total of 168 observations. Several authors have used this data set. Zeger [1988] considers a model with a latent pro-

	Forecast	α	h									
			1	2	3	4	5	6	7	8	9	10
$\lambda = 0.5$	$X_T^{(\text{Map})}(h)$	0.1	0.90	0.92	0.92	0.93	0.91	0.92	0.92	0.91	0.91	0.91
		0.5	1.13	1.27	1.33	1.35	1.36	1.35	1.35	1.37	1.39	1.37
		0.9	2.33	3.19	3.79	4.17	4.51	4.75	4.98	5.20	5.44	5.58
	$X_T^{(\text{Ei})}(h)$	0.1	0.93	0.94	0.94	0.93	0.93	0.94	0.93	0.92	0.93	0.93
		0.5	1.20	1.34	1.39	1.40	1.40	1.38	1.38	1.39	1.41	1.39
		0.9	2.36	3.21	3.83	4.22	4.58	4.82	5.06	5.28	5.52	5.65
$\lambda = 1$	$X_T^{(\text{Map})}(h)$	0.1	1.11	1.11	1.13	1.12	1.12	1.11	1.13	1.12	1.11	1.12
		0.5	1.52	1.74	1.81	1.82	1.86	1.83	1.85	1.85	1.84	1.85
		0.9	3.34	4.57	5.36	6.02	6.51	6.86	7.15	7.49	7.73	7.98
	$X_T^{(\text{Ei})}(h)$	0.1	1.24	1.23	1.25	1.23	1.24	1.23	1.25	1.25	1.24	1.25
		0.5	1.58	1.78	1.82	1.83	1.85	1.82	1.83	1.83	1.83	1.84
		0.9	3.36	4.57	5.37	6.04	6.55	6.92	7.24	7.58	7.82	8.08
$\lambda = 3$	$X_T^{(\text{Map})}(h)$	0.1	1.93	1.91	1.91	1.91	1.90	1.92	1.91	1.90	1.92	1.93
		0.5	2.59	2.93	3.01	3.05	3.04	3.08	3.13	3.10	3.12	3.10
		0.9	5.78	7.94	9.44	10.61	11.50	12.24	12.81	13.29	13.71	13.98
	$X_T^{(\text{Ei})}(h)$	0.1	1.99	1.96	1.96	1.96	1.94	1.97	1.95	1.95	1.96	1.98
		0.5	2.63	2.96	3.04	3.06	3.05	3.08	3.12	3.08	3.10	3.07
		0.9	5.79	7.95	9.44	10.62	11.52	12.27	12.86	13.36	13.79	14.07
$\lambda = 5$	$X_T^{(\text{Map})}(h)$	0.1	2.46	2.44	2.43	2.49	2.48	2.44	2.49	2.44	2.45	2.46
		0.5	3.31	3.76	3.81	3.84	3.91	3.86	3.91	3.92	3.98	3.97
		0.9	7.55	10.36	12.15	13.57	14.77	15.66	16.25	16.89	17.52	17.84
	$X_T^{(\text{Ei})}(h)$	0.1	2.50	2.48	2.46	2.52	2.51	2.47	2.53	2.47	2.49	2.50
		0.5	3.35	3.78	3.83	3.85	3.90	3.86	3.89	3.92	3.97	3.96
		0.9	7.56	10.37	12.15	13.59	14.79	15.69	16.31	16.97	17.61	17.93

Table 4.6: RMSE of $X_T^{(\text{Map})}(h)$ and $X_T^{(\text{Ei})}(h)$ for different values of α and λ considering CLS estimators for $h = 1, 2, \ldots, 10$ and sample size $T = 25$.

		α	h									
			1	2	3	4	5	6	7	8	9	10
$\lambda = 0.5$	$X_T^{(\mathrm{Map})}(h)$	0.1	0.58	0.60	0.60	0.61	0.59	0.60	0.60	0.59	0.59	0.59
		0.5	0.77	0.89	0.93	0.95	0.95	0.95	0.95	0.97	0.96	0.97
		0.9	1.65	2.30	2.74	3.01	3.23	3.41	3.58	3.74	3.90	4.02
	$X_T^{(\mathrm{Ei})}(h)$	0.1	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.55	0.55	0.55
		0.5	0.79	0.91	0.94	0.95	0.95	0.94	0.94	0.95	0.95	0.95
		0.9	1.66	2.30	2.76	3.04	3.27	3.47	3.63	3.79	3.95	4.06
$\lambda = 1$	$X_T^{(\mathrm{Map})}(h)$	0.1	0.81	0.81	0.81	0.81	0.81	0.80	0.82	0.81	0.81	0.82
		0.5	1.13	1.29	1.33	1.36	1.37	1.36	1.37	1.38	1.37	1.38
		0.9	2.52	3.46	4.07	4.56	4.91	5.14	5.37	5.61	5.80	5.94
	$X_T^{(\mathrm{Ei})}(h)$	0.1	0.88	0.88	0.89	0.88	0.89	0.88	0.89	0.89	0.89	0.89
		0.5	1.16	1.30	1.32	1.34	1.35	1.34	1.32	1.34	1.34	1.35
		0.9	2.52	3.45	4.06	4.57	4.95	5.19	5.44	5.66	5.86	6.02
$\lambda = 3$	$X_T^{(\mathrm{Map})}(h)$	0.1	1.48	1.47	1.48	1.49	1.46	1.48	1.48	1.47	1.48	1.50
		0.5	2.01	2.27	2.34	2.38	2.38	2.40	2.42	2.42	2.43	2.41
		0.9	4.52	6.20	7.38	8.29	8.96	9.57	10.01	10.36	10.67	10.90
	$X_T^{(\mathrm{Ei})}(h)$	0.1	1.51	1.49	1.49	1.50	1.48	1.50	1.49	1.49	1.49	1.52
		0.5	2.03	2.28	2.34	2.37	2.37	2.39	2.41	2.40	2.40	2.37
		0.9	4.52	6.21	7.39	8.29	8.97	9.58	10.04	10.42	10.73	10.96
$\lambda = 5$	$X_T^{(\mathrm{Map})}(h)$	0.1	1.93	1.91	1.91	1.95	1.94	1.90	1.96	1.92	1.92	1.93
		0.5	2.61	2.97	3.00	3.02	3.07	3.03	3.08	3.10	3.12	3.12
		0.9	5.96	8.20	9.55	10.67	11.64	12.35	12.80	13.26	13.78	14.02
	$X_T^{(\mathrm{Ei})}(h)$	0.1	1.94	1.93	1.92	1.96	1.97	1.91	1.97	1.93	1.93	1.94
		0.5	2.62	2.97	3.00	3.02	3.06	3.01	3.05	3.09	3.10	3.10
		0.9	5.96	8.21	9.55	10.68	11.66	12.38	12.85	13.33	13.85	14.09

Table 4.7: MAE of $X_T^{(\mathrm{Map})}(h)$ and $X_T^{(\mathrm{Ei})}(h)$ for different values of α and λ considering CLS estimators for $h = 1, 2, \ldots, 10$ and sample size $T = 25$.

		α	h									
			1	2	3	4	5	6	7	8	9	10
$\lambda = 0.5$	$X_T^{(\text{Map})}(h)$	0.1	0.90	0.92	0.91	0.92	0.92	0.93	0.91	0.93	0.92	0.93
		0.5	1.12	1.26	1.30	1.33	1.33	1.33	1.33	1.33	1.33	1.33
		0.9	2.31	3.16	3.72	4.18	4.49	4.78	5.00	5.20	5.35	5.47
	$X_T^{(\text{Ei})}(h)$	0.1	0.93	0.94	0.92	0.93	0.93	0.94	0.93	0.94	0.93	0.93
		0.5	1.21	1.34	1.37	1.39	1.38	1.38	1.38	1.38	1.37	1.36
		0.9	2.34	3.17	3.71	4.16	4.48	4.78	5.00	5.19	5.33	5.44
$\lambda = 1$	$X_T^{(\text{Map})}(h)$	0.1	1.08	1.07	1.08	1.09	1.08	1.07	1.08	1.08	1.07	1.06
		0.5	1.49	1.70	1.74	1.74	1.76	1.82	1.78	1.80	1.77	1.76
		0.9	3.28	4.49	5.28	5.91	6.34	6.74	7.00	7.22	7.44	7.64
	$X_T^{(\text{Ei})}(h)$	0.1	1.22	1.20	1.20	1.21	1.19	1.19	1.20	1.19	1.18	1.18
		0.5	1.57	1.75	1.78	1.77	1.78	1.82	1.78	1.81	1.78	1.75
		0.9	3.31	4.49	5.27	5.89	6.32	6.73	7.00	7.22	7.45	7.65
$\lambda = 3$	$X_T^{(\text{Map})}(h)$	0.1	1.87	1.89	1.88	1.90	1.87	1.90	1.89	1.89	1.89	1.90
		0.5	2.51	2.82	2.89	2.92	2.94	2.98	2.96	2.97	2.97	2.98
		0.9	5.64	7.68	9.12	10.17	10.94	11.62	12.13	12.52	12.82	13.11
	$X_T^{(\text{Ei})}(h)$	0.1	1.92	1.92	1.92	1.93	1.89	1.93	1.92	1.92	1.92	1.93
		0.5	2.57	2.85	2.92	2.94	2.95	2.99	2.96	2.97	2.97	2.98
		0.9	5.66	7.69	9.11	10.16	10.92	11.60	12.10	12.49	12.80	13.09
$\lambda = 5$	$X_T^{(\text{Map})}(h)$	0.1	2.42	2.43	2.47	2.43	2.48	2.44	2.43	2.41	2.44	2.44
		0.5	3.24	3.66	3.76	3.78	3.76	3.77	3.78	3.84	3.82	3.80
		0.9	7.31	9.86	11.80	13.17	14.13	14.93	15.66	16.09	16.59	16.90
	$X_T^{(\text{Ei})}(h)$	0.1	2.46	2.46	2.50	2.46	2.50	2.46	2.45	2.44	2.48	2.47
		0.5	3.27	3.68	3.78	3.79	3.76	3.76	3.77	3.83	3.82	3.79
		0.9	7.33	9.87	11.79	13.16	14.14	14.93	15.67	16.10	16.61	16.93

Table 4.8: RMSE of $X_T^{(\text{Map})}(h)$ and $X_T^{(\text{Ei})}(h)$ for different values of α and λ considering CLS estimators for $h = 1, 2, \ldots, 10$ and sample size $T = 50$.

		α	h									
			1	2	3	4	5	6	7	8	9	10
$\lambda = 0.5$	$X_T^{(\text{Map})}(h)$	0.1	0.57	0.59	0.58	0.58	0.58	0.59	0.58	0.60	0.59	0.60
		0.5	0.76	0.87	0.90	0.92	0.93	0.94	0.93	0.94	0.93	0.93
		0.9	1.63	2.26	2.65	2.97	3.19	3.37	3.53	3.68	3.78	3.87
	$X_T^{(\text{Ei})}(h)$	0.1	0.55	0.56	0.55	0.56	0.56	0.56	0.56	0.57	0.56	0.56
		0.5	0.79	0.89	0.92	0.93	0.94	0.93	0.94	0.94	0.92	0.92
		0.9	1.64	2.25	2.64	2.96	3.18	3.39	3.55	3.69	3.77	3.85
$\lambda = 1$	$X_T^{(\text{Map})}(h)$	0.1	0.79	0.78	0.79	0.79	0.78	0.78	0.78	0.79	0.78	0.76
		0.5	1.11	1.26	1.28	1.29	1.30	1.36	1.33	1.34	1.31	1.30
		0.9	2.46	3.37	3.96	4.42	4.74	5.03	5.22	5.39	5.58	5.71
	$X_T^{(\text{Ei})}(h)$	0.1	0.87	0.86	0.86	0.86	0.86	0.85	0.85	0.86	0.84	0.84
		0.5	1.14	1.28	1.29	1.28	1.29	1.33	1.31	1.31	1.30	1.27
		0.9	2.47	3.37	3.95	4.40	4.72	5.01	5.24	5.40	5.58	5.71
$\lambda = 3$	$X_T^{(\text{Map})}(h)$	0.1	1.45	1.46	1.45	1.46	1.44	1.47	1.45	1.45	1.45	1.46
		0.5	1.96	2.18	2.24	2.28	2.29	2.33	2.31	2.32	2.30	2.31
		0.9	4.42	6.05	7.17	7.99	8.56	9.11	9.52	9.77	10.05	10.25
	$X_T^{(\text{Ei})}(h)$	0.1	1.47	1.47	1.46	1.46	1.45	1.48	1.46	1.47	1.46	1.47
		0.5	1.99	2.20	2.25	2.28	2.29	2.32	2.30	2.31	2.29	2.30
		0.9	4.43	6.05	7.16	7.97	8.53	9.10	9.49	9.76	10.05	10.25
$\lambda = 5$	$X_T^{(\text{Map})}(h)$	0.1	1.90	1.90	1.94	1.90	1.93	1.91	1.90	1.89	1.91	1.89
		0.5	2.55	2.88	2.95	2.97	2.97	2.96	2.99	3.03	3.00	2.99
		0.9	5.77	7.77	9.32	10.34	11.15	11.79	12.38	12.67	13.10	13.33
	$X_T^{(\text{Ei})}(h)$	0.1	1.91	1.91	1.94	1.91	1.93	1.91	1.90	1.89	1.93	1.89
		0.5	2.56	2.87	2.95	2.97	2.96	2.95	2.97	3.00	2.99	2.97
		0.9	5.78	7.77	9.32	10.33	11.16	11.80	12.40	12.69	13.12	13.37

Table 4.9: MAE of $X_T^{(\text{Map})}(h)$ and $X_T^{(\text{Ei})}(h)$ for different values of α and λ considering CLS estimators for $h = 1, 2, \ldots, 10$ and sample size T = 50.

		α	h									
			1	2	3	4	5	6	7	8	9	10
$\lambda = 0.5$	$X_T^{(\mathrm{Map})}(h)$	0.1	0.91	0.92	0.91	0.93	0.92	0.91	0.93	0.92	0.92	0.92
		0.5	1.10	1.26	1.31	1.30	1.30	1.32	1.32	1.34	1.31	1.34
		0.9	2.28	3.09	3.65	4.13	4.43	4.61	4.84	4.98	5.11	5.28
	$X_T^{(\mathrm{Ei})}(h)$	0.1	0.94	0.93	0.92	0.93	0.93	0.92	0.94	0.93	0.93	0.93
		0.5	1.20	1.34	1.39	1.36	1.36	1.37	1.37	1.38	1.36	1.37
		0.9	2.31	3.11	3.64	4.10	4.39	4.55	4.77	4.89	5.00	5.14
$\lambda = 1$	$X_T^{(\mathrm{Map})}(h)$	0.1	1.08	1.05	1.07	1.06	1.06	1.08	1.07	1.06	1.05	1.07
		0.5	1.47	1.67	1.73	1.73	1.77	1.75	1.78	1.76	1.73	1.73
		0.9	3.23	4.41	5.24	5.80	6.20	6.56	6.85	7.12	7.32	7.52
	$X_T^{(\mathrm{Ei})}(h)$	0.1	1.21	1.15	1.15	1.15	1.15	1.16	1.16	1.16	1.14	1.17
		0.5	1.55	1.71	1.77	1.77	1.79	1.78	1.80	1.79	1.76	1.75
		0.9	3.27	4.42	5.23	5.77	6.15	6.51	6.78	7.03	7.22	7.42
$\lambda = 3$	$X_T^{(\mathrm{Map})}(h)$	0.1	1.86	1.86	1.89	1.88	1.89	1.89	1.87	1.85	1.87	1.88
		0.5	2.50	2.80	2.84	2.90	2.93	2.92	2.92	2.90	2.90	2.90
		0.9	5.56	7.46	8.83	9.80	10.59	11.13	11.55	11.93	12.37	12.63
	$X_T^{(\mathrm{Ei})}(h)$	0.1	1.90	1.88	1.91	1.89	1.90	1.90	1.88	1.86	1.88	1.90
		0.5	2.54	2.83	2.87	2.91	2.94	2.92	2.91	2.90	2.88	2.88
		0.9	5.58	7.46	8.83	9.80	10.59	11.12	11.53	11.91	12.35	12.60
$\lambda = 5$	$X_T^{(\mathrm{Map})}(h)$	0.1	2.41	2.42	2.43	2.43	2.40	2.42	2.40	2.44	2.41	2.45
		0.5	3.19	3.62	3.68	3.71	3.74	3.72	3.75	3.78	3.74	3.73
		0.9	7.20	9.87	11.46	12.70	13.59	14.40	15.00	15.47	15.95	16.15
	$X_T^{(\mathrm{Ei})}(h)$	0.1	2.44	2.43	2.44	2.43	2.41	2.43	2.42	2.45	2.42	2.46
		0.5	3.23	3.64	3.69	3.72	3.74	3.71	3.75	3.78	3.74	3.72
		0.9	7.22	9.86	11.46	12.69	13.58	14.38	14.98	15.44	15.91	16.11

Table 4.10: RMSE of $X_T^{(\mathrm{Map})}(h)$ and $X_T^{(\mathrm{Ei})}(h)$ for different values of α and λ considering CLS estimators for $h = 1, 2, \ldots, 10$ and sample size $T = 100$.

		α	h									
			1	2	3	4	5	6	7	8	9	10
$\lambda = 0.5$	$X_T^{(\text{Map})}(h)$	0.1	0.57	0.57	0.57	0.58	0.57	0.57	0.58	0.57	0.58	0.57
		0.5	0.74	0.86	0.90	0.90	0.90	0.91	0.91	0.93	0.92	0.93
		0.9	1.61	2.19	2.61	2.94	3.15	3.29	3.44	3.53	3.62	3.74
	$X_T^{(\text{Ei})}(h)$	0.1	0.56	0.55	0.55	0.56	0.55	0.55	0.56	0.55	0.56	0.56
		0.5	0.79	0.89	0.93	0.92	0.92	0.93	0.92	0.94	0.92	0.94
		0.9	1.60	2.19	2.59	2.91	3.12	3.25	3.41	3.49	3.56	3.66
$\lambda = 1$	$X_T^{(\text{Map})}(h)$	0.1	0.78	0.76	0.77	0.77	0.78	0.78	0.78	0.78	0.76	0.78
		0.5	1.08	1.23	1.27	1.28	1.30	1.30	1.31	1.30	1.30	1.29
		0.9	2.42	3.30	3.91	4.34	4.64	4.91	5.12	5.33	5.48	5.61
	$X_T^{(\text{Ei})}(h)$	0.1	0.86	0.82	0.82	0.83	0.84	0.84	0.83	0.84	0.82	0.84
		0.5	1.13	1.24	1.28	1.28	1.29	1.30	1.32	1.30	1.30	1.28
		0.9	2.44	3.29	3.88	4.31	4.62	4.89	5.09	5.28	5.43	5.56
$\lambda = 3$	$X_T^{(\text{Map})}(h)$	0.1	1.44	1.43	1.45	1.43	1.45	1.45	1.43	1.43	1.44	1.44
		0.5	1.94	2.17	2.21	2.27	2.27	2.28	2.27	2.26	2.25	2.26
		0.9	4.34	5.85	6.92	7.64	8.26	8.65	9.00	9.26	9.62	9.82
	$X_T^{(\text{Ei})}(h)$	0.1	1.45	1.43	1.46	1.44	1.46	1.45	1.43	1.43	1.44	1.45
		0.5	1.96	2.18	2.23	2.26	2.26	2.26	2.25	2.24	2.22	2.23
		0.9	4.35	5.85	6.93	7.64	8.27	8.65	9.00	9.26	9.63	9.83
$\lambda = 5$	$X_T^{(\text{Map})}(h)$	0.1	1.90	1.88	1.91	1.89	1.88	1.89	1.88	1.91	1.88	1.92
		0.5	2.51	2.85	2.89	2.91	2.96	2.94	2.95	2.97	2.95	2.94
		0.9	5.67	7.80	9.02	10.02	10.73	11.33	11.81	12.18	12.60	12.76
	$X_T^{(\text{Ei})}(h)$	0.1	1.90	1.87	1.91	1.87	1.88	1.88	1.88	1.90	1.87	1.91
		0.5	2.52	2.86	2.89	2.90	2.94	2.92	2.94	2.96	2.93	2.92
		0.9	5.68	7.79	9.02	10.01	10.71	11.31	11.79	12.17	12.58	12.74

Table 4.11: MAE of $X_T^{(\text{Map})}(h)$ and $X_T^{(\text{Ei})}(h)$ for different values of α and λ considering CLS estimators for $h = 1, 2, \ldots, 10$ and sample size $T = 100$.

cess that generates overdispersion and autocorrelation. Brännäs and Johansson [1994] study properties of different estimators for the parameters related to the latent variable in the model proposed by Zeger [1988]. Jorgensen et al. [1999] propose a nonstationary state space model. Davis et al. [2000] propose a approach to diagnosing the existence of a latent stochastic process in the mean of a Poisson regression model and provide formulae for the effect of autocovariance on standard errors of the regression coefficients. Heinen [2003] proposed the ACP model and Silva [2005] studied different estimation methods and a criterion for order selection in INAR model. All these researchers used the data Polio as an application of their investigations. Since the data Polio is overdispersed, Heinen [2003] and Silva [2005] remark that it is not correct to assume the Poisson marginal distribution, but it is more appropriate to think of an INARCH(1) model taking into account the overdispersion.

The second application considered the monthly strike data published by the U.S. Bureau of Labor Statistics for the period between January 1994 until December 2002, a total of 108 observations. The counts describe the number of work stoppages leading to 1,000 workers or more being idle in effect in the period. This data set was used initially by Jung et al. [2005] who fitted a Poisson INAR(1) model to the data, but since the estimates obtained with different methods deviated heavily from each other, they concluded that such a model is not appropriate. In fact the data set exhibits overdispersion, making the Poisson marginal distribution an unreasonable choice. Weiß [2010] showed that the data set is modeled very well by an INARCH(1) model.

We refer to the two data sets as POLIO and STRIKE respectively. Clearly the number of new polio cases at month t, X_t, can be viewed as the sum of the number of cases generated (by contagious) from infected people at month $t-1$, $\alpha * X_{t-1}$, and the immigration cases, i.e., infected people that arrived to U.S. between months $t-1$ and t, ϵ_t. On the other hand, the observed number of work stoppages leading to 1,000 workers or more being idle in effect at any month t, X_t, can be viewed as the sum of the number of work stoppages leading to 1,000 workers or more being idle in effect at month $t-1$ and continue on work stoppage, $\alpha * X_{t-1}$, and the number of newly work stoppages leading to 1,000 workers or more being idle in effect that were started between the months $t-1$ and t, ϵ_t.

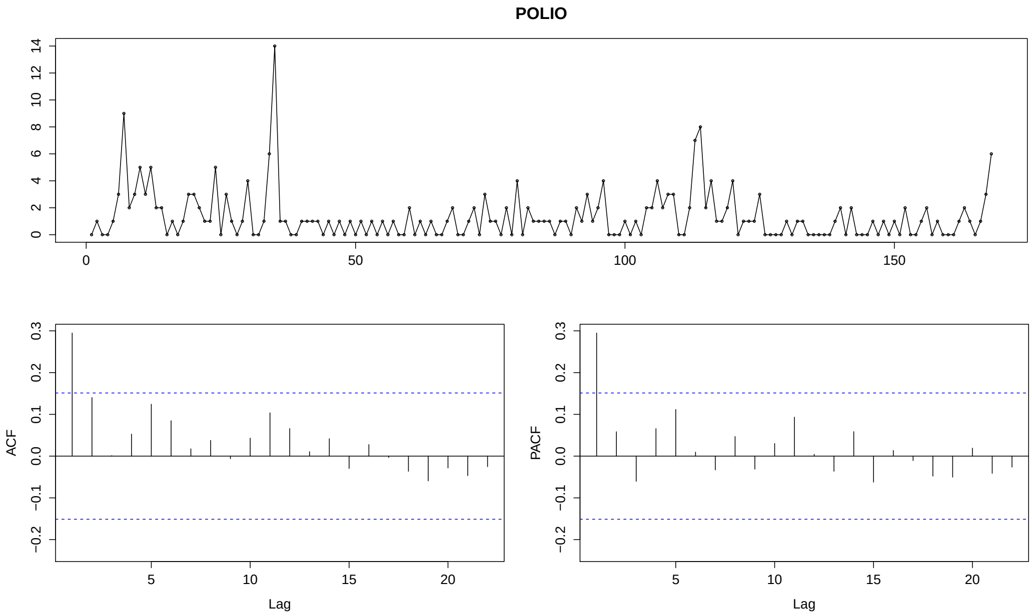

Figure 4.1: Monthly counts of POLIO data, January 1970-December 1983 and sample autocorrelation and partial autocorrelation functions.

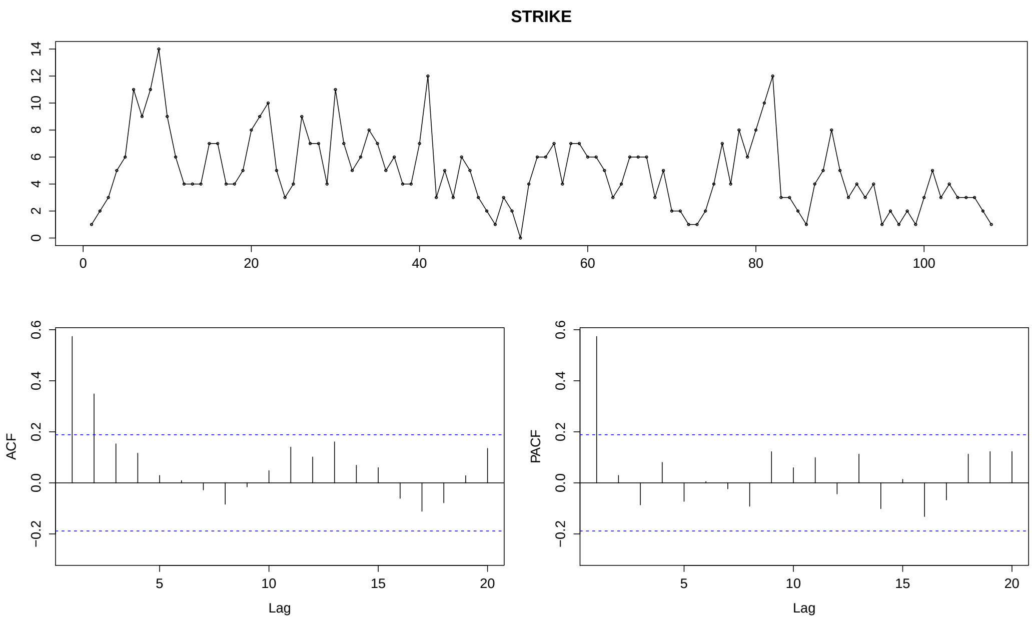

Figure 4.2: Monthly counts of STRIKE data, January 1994-December 2002 and sample autocorrelation and partial autocorrelation functions.

Data	Minimum Count	Maximum Count	Median	Mean	Variance
POLIO	0	14	1	1.333	3.484
STRIKE	0	14	4	4.944	7.849

Table 4.12: Summary Statistics for POLIO and STRIKE data

A summary of their simple descriptive statistics is reported in Table 4.12. Note that for both data sets the variance is greater than the mean, i. e., the data sets shows overdispersion. Then, the Poisson INAR(1) process is an unreasonable selection.

The sample autocorrelation and sample partial autocorrelation functions of the polio data are shown in Figure 4.1. The analysis of these figures suggests a first order autoregressive process and this process should take into account the overdispersion. Then, it is reasonable to think of an INARCH(1) process for the POLIO data. Figure 4.2 provides the time series plots of STRIKE data as well as their corresponding sample autocorrelation and sample partial autocorrelation functions. From the partial autocorrelation function, it becomes clear that a first order autoregressive process seems to be reasonable, and given that the data set presents overdispersion the INARCH(1) model is a good choice to fit the STRIKE data.

		CLS		CML			CLS			CML		
Month/Year	T	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\lambda}$	x_{T+1}	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Map})}(1)$	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Map})}(1)$	$x_T^{(\mathrm{Md})}(1)$
May/82	148	0.293	0.978	0.353	0.896	0	1	1	1	1	1	1
Jun/82	149	0.294	0.968	0.353	0.888	1	1	1	0	1	1	0
Jul/82	150	0.294	0.969	0.352	0.889	0	1	1	1	1	1	1
Aug/82	151	0.295	0.959	0.351	0.882	2	1	1	0	1	1	0
Set/82	152	0.292	0.969	0.346	0.896	0	1	1	1	1	1	1
Oct/82	153	0.29	0.962	0.341	0.893	0	1	1	0	1	1	0
Nov/82	154	0.293	0.952	0.345	0.882	1	1	1	0	1	1	0
Dec/82	155	0.293	0.953	0.345	0.883	2	1	1	1	1	1	1
Jan/83	156	0.292	0.958	0.345	0.888	0	1	1	1	1	1	1
Feb/83	157	0.29	0.951	0.34	0.884	1	1	1	0	1	1	0
Mar/83	158	0.29	0.951	0.339	0.886	0	1	1	1	1	1	1
Apr/83	159	0.291	0.942	0.339	0.879	0	1	1	0	1	1	0
May/83	160	0.293	0.933	0.342	0.869	0	1	1	0	1	1	0
Jun/83	161	0.295	0.925	0.346	0.858	1	1	1	0	1	1	0
Jul/83	162	0.295	0.925	0.345	0.86	2	1	1	1	1	1	1
Aug/83	163	0.295	0.931	0.346	0.864	1	1	1	1	1	1	1
Set/83	164	0.294	0.928	0.344	0.863	0	1	1	1	1	1	1
Oct/83	165	0.295	0.92	0.343	0.857	1	1	1	0	1	1	0
Nov/83	166	0.295	0.921	0.343	0.858	3	1	1	1	1	1	1
Dec/83	167	0.294	0.933	0.344	0.868	6	2	2	1	2	2	1
-	-	-	-	-	-	MSE	1.6	1.6	2.3	1.6	1.6	2.3
-	-	-	-	-	-	MAE	0.9	0.9	1.1	0.9	0.9	1.1

Table 4.13: Point prediction one-step ahead of monthly count May 1982 to December 1983 for POLIO data.

		CLS		CML			CLS			CML		
Month/Year	T	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\lambda}$	x_{T+1}	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Map})}(1)$	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Md})}(1)$	$x_T^{(\mathrm{Map})}(1)$	$x_T^{(\mathrm{Md})}(1)$
May/2001	88	0.534	2.554	0.594	2.232	8	5	5	5	5	5	5
Jun/2001	89	0.532	2.594	0.595	2.254	5	7	7	6	7	7	7
Jul/2001	90	0.525	2.61	0.587	2.276	3	5	5	5	5	5	5
Aug/2001	91	0.527	2.578	0.586	2.258	4	4	4	4	4	4	4
Set/2001	92	0.527	2.573	0.586	2.258	3	5	5	4	4	4	4
Oct/2001	93	0.53	2.538	0.588	2.231	4	4	4	4	4	4	3
Nov/2001	94	0.531	2.535	0.588	2.231	1	4	4	4	4	4	4
Dec/2001	95	0.537	2.46	0.591	2.175	2	3	3	2	3	3	2
Jan/2002	96	0.543	2.42	0.599	2.127	1	3	3	3	3	3	3
Feb/2002	97	0.554	2.337	0.612	2.036	2	3	3	2	2	2	2
Mar/2002	98	0.559	2.303	0.617	1.998	1	3	3	3	3	3	3
Apr/2002	99	0.569	2.228	0.629	1.916	3	3	3	2	2	2	2
May/2002	100	0.567	2.235	0.625	1.941	5	4	4	3	4	4	3
Jun/2002	101	0.565	2.26	0.622	1.966	3	5	5	5	5	5	5
Jul/2002	102	0.565	2.238	0.62	1.957	4	4	4	3	4	4	3
Aug/2002	103	0.565	2.24	0.62	1.961	3	4	4	4	4	4	4
Set/2002	104	0.567	2.215	0.62	1.945	3	4	4	3	4	4	3
Oct/2002	105	0.569	2.195	0.622	1.928	3	4	4	3	4	4	3
Nov/2002	106	0.571	2.175	0.623	1.913	2	4	4	3	4	4	3
Dec/2002	107	0.576	2.134	0.627	1.878	1	3	3	3	3	3	3
	-	-	-	-	-	MSE	2.8	2.8	2.4	2.65	2.65	2.6
	-	-	-	-	-	MAE	1.4	1.4	1.2	1.35	1.35	1.3

Table 4.14: Point prediction one-step ahead of monthly count May 2001 to December 2002 for STRIKE data.

In order to compare the different forecasts one-step ahead, we found the one-step ahead point prediction of the monthly number of Polio cases in United States from May 1982 to December 1983 as well as of the monthly number of strikes from May 2001 to December 2002. Table 4.13 and Table 4.14 present the point forecasts based on the exact median, approximate median and mode of the Poisson distribution with parameter $\alpha X_T + \lambda$ as well as the CLS and CML estimates for α and λ for POLIO and STRIKE data, respectively.

For POLIO data, the conditional median $X_T^{(\mathrm{Md})}(1)$ and its approximation $X_T^{(\mathrm{Map})}(1)$ produce the same MSE and MAE, and they are lower than the MSE and MAE of the conditional mode; also, the CLS and CML estimation methods produce the same results.

For STRIKE data, the MSE and MAE of the conditional mode $X_T^{(\mathrm{Md})}(1)$ is slightly lower than the MSE and MAE of $X_T^{(\mathrm{Md})}(1)$ and its approximation $X_T^{(\mathrm{Map})}(1)$. Besides, the CLS and CML produce almost the same results.

Furthermore, note that for the POLIO data

$$\text{MSE of } X_T^{(\mathrm{Md})}(1) \text{ / MSE of } X_T^{(\mathrm{Md})}(1) = 1.6/2.3 \approx 0.70$$

and for the STRIKE data

$$\text{MSE of } X_T^{(\mathrm{Md})}(1) \text{ / MSE of } X_T^{(\mathrm{Md})}(1) = 2.4/2.8 \approx 0.86,$$

so, in the data set where the conditional median and its approximation are better, they produce a reduction of 30% of the MSE while in the data set where the conditional mode is better, the conditional mode produces only a reduction of 14% of the MSE.

In order to compare the different forecasts two-steps ahead, we found the two-steps ahead point prediction of the monthly number of strikes from May 2001 to December 2002. Table 4.15 presents the point forecasts based on the exact median $X_T^{(\mathrm{Mn})}(2)$, the integer part of the mean $X_T^{(\mathrm{Ei})}(2)$, and on the mode $X_T^{(\mathrm{Md})}(2)$ of the distribution of X_{T+2} given X_T as well as the recursive forecast $X_T^{(\mathrm{Map})}(2)$ proposed in equation (4.37) with $h = 2$. The CLS and CML estimates for α and λ are also included.

Note that, CML estimators are slightly better than CLS estimators in terms of MSE and MAE. Also, using CLS estimates $X_T^{(\mathrm{Md})}(2)$ is better than the others forecast, being $X_T^{(\mathrm{Map})}(2)$ the second best and $X_T^{(\mathrm{Ei})}(2)$ the worst, in terms of MSE and MAE. Using CML estimates $X_T^{(\mathrm{Md})}(2)$ is better than the other forecasts, and they are competitive in terms of MSE and MAE.

In Figure 4.3 and Figure 4.4, we present the one-step ahead forecasts using a 95% upper prediction interval for POLIO and STRIKE data sets. Notice that for POLIO data, seven observations fell outside of the interval, while for STRIKE data only two observations fell outside of the interval. This fact may possibly be explained by interventions, i. e., some unusual fact that happened in June/96 and May/97 for STRIKE data and in the months of Dec/71, Oct/72, Nov/72, Aug/76, May/79, Jun/79 and Dec/83 for POLIO data.

		CLS		CML			CLS				CML			
Month/Year	T	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\lambda}$	x_{T+2}	$x_T^{(\mathrm{Mn})}(2)$	$x_T^{(\mathrm{Map})}(2)$	$x_T^{(\mathrm{Md})}(2)$	$x_T^{(\mathrm{Ei})}(2)$	$x_T^{(\mathrm{Mn})}(2)$	$x_T^{(\mathrm{Map})}(2)$	$x_T^{(\mathrm{Md})}(2)$	$x_T^{(\mathrm{Ei})}(2)$
May/2001	87	0.534	2.548	0.594	2.225	8	5	5	4	5	5	4	4	4
Jun/2001	88	0.534	2.554	0.594	2.232	5	5	5	5	5	5	5	4	5
Jul/2001	89	0.532	2.594	0.595	2.254	3	6	7	5	6	6	7	6	6
Aug/2001	90	0.525	2.61	0.587	2.276	4	5	5	5	5	5	5	5	5
Set/2001	91	0.527	2.578	0.586	2.258	3	5	4	4	4	4	4	4	4
Oct/2001	92	0.527	2.573	0.586	2.258	4	5	5	4	5	5	4	4	4
Nov/2001	93	0.53	2.538	0.588	2.231	1	4	4	4	4	4	4	4	4
Dec/2001	94	0.531	2.535	0.588	2.231	2	5	4	4	5	5	4	4	4
Jan/2002	95	0.537	2.46	0.591	2.175	1	4	3	3	4	4	3	3	3
Feb/2002	96	0.543	2.42	0.599	2.127	2	4	3	4	4	4	3	3	4
Mar/2002	97	0.554	2.337	0.612	2.036	1	4	3	3	3	3	2	3	3
Apr/2002	98	0.559	2.303	0.617	1.998	3	4	3	3	4	4	3	3	3
May/2002	99	0.569	2.228	0.629	1.916	5	4	3	3	3	3	2	3	3
Jun/2002	100	0.567	2.235	0.625	1.941	3	4	4	4	4	4	4	3	4
Jul/2002	101	0.565	2.26	0.622	1.966	4	5	5	4	5	5	5	4	5
Aug/2002	102	0.565	2.238	0.62	1.957	3	4	4	4	4	4	4	3	4
Set/2002	103	0.565	2.24	0.62	1.961	3	5	4	4	4	4	4	4	4
Oct/2002	104	0.567	2.215	0.62	1.945	3	4	4	4	4	4	4	3	4
Nov/2002	105	0.569	2.195	0.622	1.928	2	4	4	4	4	4	4	3	4
Dec/2002	106	0.571	2.175	0.623	1.913	1	4	4	4	4	4	4	3	4
	-	-	-	-	-	MSE	4.35	3.6	3.4	3.95	3.95	4	3	3.7
	-	-	-	-	-	MAE	1.85	1.6	1.5	1.75	1.75	1.6	1.3	1.6

Table 4.15: Point prediction two-steps ahead of monthly count May 2001 to December 2002 for STRIKE data.

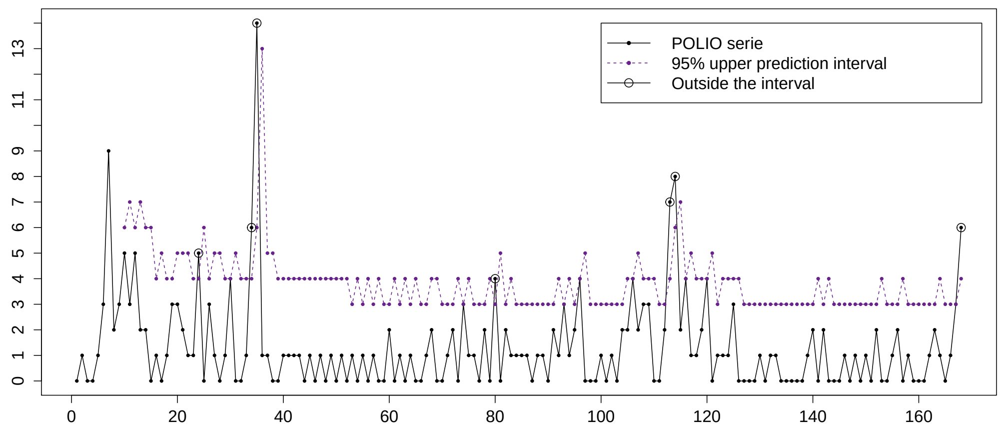

Figure 4.3: Monthly counts of POLIO data, January 1970-December 1983 and one-step ahead forecasts confidence limits.

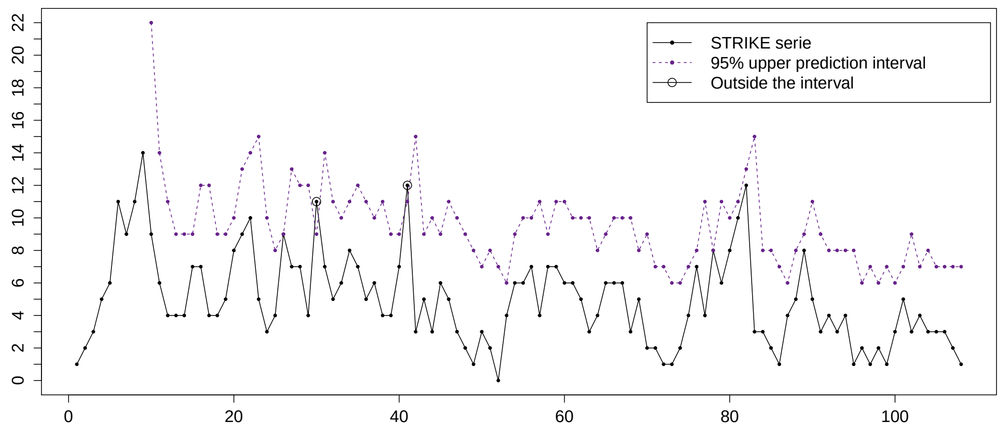

Figure 4.4: Monthly counts of STRIKE data, January 1994-December 2002 and one-step ahead forecasts confidence limits.

Chapter 5

Main contributions and future work

Resumo

Neste capítulo apresentamos as principais contribuições para a previsão dos processos autorregressivos de primeira ordem Poisson INAR(1), estudado no capítulo 3, e INARCH(1) estudado no capítulo 4. Além disso, apresentamos os tópicos de pesquisa futura.

5.1 Main contributions

The aim of this thesis is to provide some contributions to the forecasting for the Poisson INAR(1) and the INARCH(1) processes. A literature review is given with the purpose of presenting the existing methodologies and to provide a solid base for the construction of new methodologies. For these processes the predictors of a future value must be an integer value. Some researchers have studied how to produce coherent forecasts for the INAR(1) process, i.e., integer-valued forecasts, however to the best of our knowledge forecasting in the INARCH(1) process has not been studied.

Chapter 3 studied forecasting in INAR(1) processes. The main contributions of this chapter can be summarized as follows:

✓ It were provided the analytical expression for the expected value of the integer part of the one-step ahead conditional mean.

✓ It was studied by Monte Carlo simulation the behavior of three forecasts: the integer part of the conditional mean, the conditional median and the conditional mode. For known parameters it can be concluded that the three forecasts were competitive, the integer part of conditional mean being a little worse for $\alpha = 0.9$.

✓ For unknown parameters it was used YW, CLS and CML estimation methods, in this case it can be concluded that for large values of α and small sample sizes YW estimators were worse than CLS and CML estimators, CLS being slightly better than CML, in terms of RMSE and MAE. For small and moderate sample sizes the conditional mean was slightly better than the others. For large sample sizes the three forecasts: conditional median, conditional mode and integer part of the conditional mean, and, the three estimation methods considered were competitive, in terms of RMSE and MAE.

✓ It was suggested to use the conditional median as forecast and CLS estimators, because they have explicit expression while CML are calculated using numerical maximization and CML estimators are only a little better than CLS estimators, in terms, of RMSE and MAE.

✓ It was studied the predictive power of Poisson INAR(1) model under misspecified data. The arrival process was misspecified by letting its distribution be uniform over $\{0, 1, 2, 3\}$. From the Monte Carlo simulation study it can be deduced that the predictive power does not deteriorate even when this misspecified model is used.

✓ It was compared the one-step ahead forecasts in two data sets. The first data set relates to claimants who have had soft tissue injures, such as contusions and bruises, while the second data set relates to claimants with dislocations. The first and second data sets are referred as SOFT INJURES and DISLOCATIONS respectively. For SOFT INJURES data the integer part of conditional mean is a little better than the other forecasts in terms of MSE and MAE. For DISLOCATIONS data using CLS estimates, the conditional median is better than the conditional mode and the integer part of the conditional mean, while using CML estimates the conditional median and mode are competitive and they are better than the integer part of conditional mean in terms of MSE and MAE. However, in the data set where the conditional median and mode are better they produce a reduction of 46% of the MSE, while in the data set where the integer part of the conditional mean is better it produces only a reduction of 7% of the MSE. Then, it can be concluded that the conditional median is in general preferable to the other two forecasts.

In Chapter 4 it was presented a different way to define the INARCH(1) process based on the Poisson thinning operator and it was studied the forecasting in this process. The main contributions of this chapter can be summarized as follows.

- ✓ The INARCH(1) process was defined based on the Poisson thinning operator. Also, several properties of the Poisson thinning operator were found and proved .
- ✓ By using an easy argument of Markov chain, it was demonstrated that the marginal stationary distribution of the INARCH(1) process exists and it is unique.
- ✓ It were found analytic expressions for the $r-$th marginal ordinary moment and for the $h-$steps conditional mean and variance. Also, it was provided an expression for the $h-$steps ahead conditional probability generating function. For $h = 2$, it was found a simple expression for the conditional probability function.
- ✓ Given that the one-step ahead distribution is a Poisson distribution, its median and mode were proposed as one-step ahead forecasts; however, given that there is no closed expression for the median of a Poisson distribution, it was provided a coherent approximation and its mean and variance limits were found. Also, it was proposed a recursive expression for the $h-$steps ahead forecast, for $h \geqslant 2$.
- ✓ For the proposed approximate median it was demonstrated weakly conditional consistency and then it was demonstrated strongly conditional consistency. Moreover, it was proved weakly consistency and then it was demonstrated strongly consistency.

✓ It was studied by Monte Carlo simulation the behavior of the one-step ahead proposed forecasts: conditional mode, conditional median and its approximation. For known and unknown parameters it can be concluded that the conditional median and its approximation provide almost the same results and they are slightly better than the conditional mode both in terms of RMSE and in terms of MAE. Additionally, for unknown parameters YW estimators were a little worse than the others for small and moderated sample sizes, in terms of RSME and MAE. For large T the estimation methods and the forecasts considered provided almost the same results, in terms of RMSE and MAE. Therefore we suggest to use the approximated conditional median because it has an easy analytical expression and was slightly better in terms of RMSE and MAE.

✓ Also, by Monte Carlo simulation it was studied the behavior of the two-steps ahead forecasts: exact conditional mean, conditional mode, the recursive forecast proposed considering $h = 2$ and the integer part of the two-steps ahead conditional mean. It can be concluded that the conditional mode was the worst forecast, the conditional median and the recursive forecast provided almost the same results and for large T the conditional median, the recursive forecast and the integer part of the conditional mean were competitive, in terms of RMSE and MAE. CLS and CML estimators provided almost the same results, CML estimators being slightly better than CLS estimators, in terms of RMSE and MAE. Therefore, we suggest to use the recursive forecast because in general it was slightly better than the others forecasts.

✓ For one and two-steps ahead forecasts we suggest to use CLS estimators. The reason for this is that they produce almost the same results than CML estimators, CML being slightly better than CLS, in terms of RMSE and MAE, but CLS estimators have the advantage of being found more easily from explicit formulas while CML estimators are found by complicated numerical maximization.

✓ It was studied by Monte Carlo simulation the behavior of the proposed recursive forecast and the integer part of conditional mean for $h = 1, 2, \ldots, 10$. The predictors have similar behavior in terms of RMSE and MAE and we were not able to conclude that none is better than the other. Further, we found that for small

α the RMSE and the MAE of the forecast one-step ahead are not much smaller than those of the forecasts $h = 2, \ldots, 10$ steps ahead, while for large α the RMSE and the MAE increase considerably from one-step ahead to two-steps ahead and then increases more slowly, for both predictors. For the proposed recursive forecast, this can be explained from the fact that the forecasts one-step ahead use the conditional distribution, but the forecasts $h \geqslant 2$ steps ahead only use a recursive expression without using the conditional distribution h steps ahead.

✓ It were used two data sets to illustrate and compare the one and two-steps ahead forecasts. The first application considers the monthly number of Polio cases in the United States, referred as POLIO, while the second application considers the monthly strike data published by the U.S. Bureau of Labor Statistics, referred as STRIKE. For POLIO data the one-step ahead conditional median and its approximation are better than the one-step ahead conditional mode, in terms of MSE and MAE, while for STRIKE data the one-step ahead conditional mode is slightly better than the one-step ahead conditional median and its approximation, in terms of MSE and MAE. However, in the data set where the conditional median and its approximation are better, they produce a reduction of 30% of the MSE while in the data set where the conditional mode is better, the conditional mode produces only a reduction of 14% of the MSE. On the other hand, for STRIKE data and using the CLS estimates, the two-steps ahead recursive forecast and the two-steps ahead conditional mode are competitive and better than the two-steps ahead conditional median, in terms of MSE and MAE. However, using CML estimates, the two-steps ahead recursive forecast and the two-steps ahead conditional median are competitive and better than the two-steps ahead conditional mode. Therefore, we suggest to use the two-steps ahead recursive forecast as two-steps ahead predictor. Finally using the 95% upper prediction interval proposed, it was found that for POLIO data seven observations fell outside of the interval, while for STRIKE data only two observations fell outside of the interval. This fact may possibly be explained by interventions, i. e., some unusual fact that happened in June/96 and May/97 for STRIKE data and in the months of Dec/71, Oct/72, Nov/72, Aug/76, May/79, Jun/79 and Dec/83 for POLIO data.

5.2 Future work

5.2.1 Additional properties of proposed forecasts

An immediate extension of this work is to study additional properties of

$$X_T^{(\mathrm{Map})}(1) = \left\lceil \hat{\alpha} X_T + \hat{\lambda} - \frac{2}{3} \right\rceil$$

as predictor of X_{T+1} given X_T. It could be interesting to study properties for the moments of

$$X_T^{(\mathrm{Map})}(h) = \left\lceil \hat{\alpha}\, \hat{\tilde{X}}_{T+h-1} + \hat{\lambda} - \frac{2}{3} \right\rceil,$$

the $h-$steps aheadestimator as predictor of X_{T+h} given X_T, for $h \geqslant 2$. So, maybe it could be possible to prove consistency of the $h-$steps aheadproposed predictors.

5.2.2 Forecasting for p order processes

An immediate extension of this research could be to consider the forecasting on INAR(p) or INARCH(p) processes. A discrete non-negative integer-valued process $(X_t)_{t\geqslant 1}$, $X_t \in \mathbb{N}_0$, is called an INAR(p) process if it satisfies the recursive equation

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \cdots + \alpha_p \circ X_{t-p} + \epsilon_t \quad \text{for } t \geqslant 1,$$

where $\{\alpha_1, \ldots, \alpha_p\} \subset (0,1)$, $\alpha_1 + \cdots + \alpha_p < 1$, '$\circ$' represents the binomial thinning operator introduced in Definition 1, $(\epsilon_t)_{t\geqslant 1}$ is a sequence of non-negative integer-valued i.i.d. random variables with mean μ_ϵ and variance σ_ϵ^2. It is assumed that all counting series of $\alpha_i \circ X_t$, $i = 1, 2, \ldots, p$ are independent of each other and independent of ϵ_t. Thus, $\alpha_1 \circ X_{t-1}, \ldots, \alpha_p \circ X_{t-p}$ given $X_{t-1}, \ldots, X_{t-p}$ are binomial independent random variables. On the other hand, we can consider the INARCH(p) process. A discrete non-negative integer-valued process $(X_t)_{t\geqslant 1}$, $X_t \in \mathbb{N}_0$, is called an INARCH(p) process if it satisfies the recursive equation

$$X_t = \alpha_1 * X_{t-1} + \cdots + \alpha_p * X_{t-p} + \epsilon_t \quad \text{for } t \geqslant 1,$$

where $\{\alpha_1, \ldots, \alpha_p\} \subset (0,1)$, $\alpha_1 + \cdots + \alpha_p < 1$, '$\circ$' represents the Poisson thinning operator given in Definition 4.1, $(\epsilon_t)_{t\geqslant 1}$ is a sequence of non-negative integer-valued i.i.d. random variables with mean μ_ϵ and variance σ_ϵ^2. It is assumed that the all counting series of $\alpha_i \circ X_t$, $i = 1, 2, \ldots, p$ are independent of each other and independent of ϵ_t. Thus, $\alpha_1 * X_{t-1}, \ldots, \alpha_p \circ X_{t-p}$ given $X_{t-1}, \ldots, X_{t-p}$ are Poisson independent random variables.

The INAR(p) process has an interesting interpretation: consider a cell population in which each cell can survive until time p, i. e., they can survive up to p generations. Thus, if the probability of surviving during exactly k generations is α_k, then $\alpha_k \circ X_{t-k}$ represents the number of individuals in the population in time $t-k$ that will survive up to time t. Therefore, the population size X_t at time t is $S_1 + \cdots + S_p + \epsilon_t$, where S_k is the number of individuals surviving during exactly k generations, this is $\alpha_k \circ X_{t-k}$, plus the total number of immigrants ϵ_t. Note that at each time the fraction of the total population surviving at least to the next generation is $\beta \circ X_t$, where $\beta = \alpha_1 + \cdots + \alpha_p$. Also, if ϵ_t has Poisson distribution with mean λ is easy to prove that the stationary distribution of X_t is Poisson with mean $\lambda/(1-\beta)$.

The INARCH(p) process can be interpreted similarly if we consider that each cell can divide up to the $p-$th generation. Although, in this process the stationary distribution is not trivial, the conditional distribution is simple; if ϵ_t has Poisson distribution with mean λ, then the conditional distribution of X_t given all the past of the series is Poisson with mean $\alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \lambda$.

Therefore, an interesting future work could be to develop a study investigating the performance of the proposed estimators and predictors for higher order processes analogous to the study that was done here for the first-order processes. Another possible research topic is the behavior of the forecasts when $\beta = \alpha_1 + \cdots + \alpha_p = 1$, i. e., when there is a root of the autoregressive polynomial at the boundary of the unit circle. Also, it could be studied how forecasts can be affected if the process order is misspecified.

5.2.3 Forecasting for signed processes

The topic that has been attracted much attention among researchers is the study of the signed integer-value processes. If X is a integer-value random variable, not necessarily with positive values, and if $\alpha \in (0,1)$, it can be defined $\alpha \circ X = (\operatorname{sgn} X)(\alpha \circ |X|)$. This is, for $X < 0$ is defined $\alpha \circ X = -[\alpha \circ (-X)]$. Notice that $\alpha \circ X$ can be interpreted as thinning operator, in the sense of $|\alpha \circ X| \leqslant |X|$.

Thus, it can be defined the signed INAR(1) process by the recursion

$$X_t = \alpha \circ X_{t-1} + \epsilon_t \quad \text{para} \;\; t \geqslant 1,$$

where it is assumed that the support of the ϵ_t distribution is the positive and negative integers. For example, a signed Poisson distribution can be considered as ϵ_t distribution. This is, is supposed that ϵ_t has symmetry distribution around zero and the distribution of $|\epsilon_t|$ is a Poisson distribution.

It could be studied what properties the stationary distribution must be have. Also it can be investigated the properties of the estimators for unknown parameters and how to use them to predict future values.

Appendix

In this Appendix we present the proofs of properties of Poisson thinning operation exhibited in lemma 2

i) $0 * X_1 = \sum_{i=0}^{X_1} N_i$, where $N_i \sim$ Poisson(0), thus P(N_i = 0) = 1 and $0 * X_1 = \sum_{i=0}^{X_1} 0 = 0$.

ii) Let $\mathcal{G}_Z(s)$ be the probability generated function (PGF) of random variable Z defined by $\mathcal{G}_Z(s) = \mathrm{E}\left[s^Z\right]$, we have

$$
\begin{aligned}
\mathcal{G}_{\alpha_1 * (X_1 + X_2)}(s) = &\ \mathrm{E}\left[s^{\alpha_1 * (X_1 + X_2)}\right] \\
&= \mathrm{E}\left[s^{\sum_{i=1}^{X_1+X_2} N_i}\right], \text{ where the } N_i' \text{ s are i.i.d with} N_i \sim \text{ Poisson}(\alpha_1) \\
&= \mathrm{E}\left[s^{\sum_{i=1}^{X_1} N_i + \sum_{j=X_1+1}^{X_1+X_2} N_j}\right] \\
&= \mathrm{E}\left[s^{\sum_{i=1}^{X_1} N_i}\ s^{\sum_{j=1}^{X_2} N_j}\right], \text{ where } N_j'\text{s are independent of } N_i'\text{s} \\
&= \mathrm{E}\left[s^{\sum_{i=1}^{X_1} N_i}\right] \mathrm{E}\left[s^{\sum_{j=1}^{X_2} N_j}\right] \\
&= \mathrm{E}\left[s^{\alpha_1 * X_1}\right] \mathrm{E}\left[s^{\alpha_1 * X_2}\right] \\
&= \mathcal{G}_{\alpha_1 * X_1}(s)\, \mathcal{G}_{\alpha_1 * X_2}(s) \\
&= \mathcal{G}_{\alpha_1 * X_1 + \alpha_1 * X_2}(s).
\end{aligned}
$$

Then, using the uniqueness of PFG we have proved the result.

iii) For the next demonstrations we use the fact

$$\alpha_1 * Z\,|Z \sim \text{Poisson}(\alpha_1 Z).$$

Thus, we have

$\mathrm{E}[\alpha_1 * X_1] = \mathrm{E}\{\mathrm{E}[\alpha_1 * X_1 | X_1]\} = \mathrm{E}\{\mathrm{E}[\text{Poisson}(\alpha_1 X_1)]\} = \mathrm{E}[\alpha_1 X_1] = \alpha_1 \mathrm{E}[X_1]$.

iv)

$$\begin{aligned}
\mathrm{Var}[\alpha_1 * X_1] &= \mathrm{E}\big[\mathrm{Var}(\alpha_1 * X_1 | X_1)\big] + \mathrm{Var}\big[\mathrm{E}(\alpha_1 * X_1 | X_1)\big] \\
&= \mathrm{E}\big[\mathrm{Var}(\text{Poisson}(\alpha_1 X_1))\big] + \mathrm{Var}\big[\mathrm{E}(\text{Poisson}(\alpha_1 X_1))\big] \\
&= \mathrm{E}\big[\alpha_1 X_1\big] + \mathrm{Var}\big[\alpha_1 X_1\big] \\
&= \alpha_1 \mathrm{E}\big[X_1\big] + \alpha_1^2 \mathrm{Var}\big[X_1\big].
\end{aligned}$$

Properties *v)* and *vi)* follow from the definition.

vii)

$$\begin{aligned}
\mathrm{Cov}(\alpha_1 * X_1, \alpha_2 * X_2) &= \mathrm{E}\big[\mathrm{Cov}\big(\alpha_1 * X_1, \alpha_2 * X_2 | \sigma(X_1, X_2)\big)\big] \\
&\quad + \mathrm{Cov}\big[\mathrm{E}(\alpha_1 * X_1 | \sigma(X_1, X_2)), \mathrm{E}(\alpha_2 * X_2 | \sigma(X_1, X_2))\big] \\
&= \mathrm{E}\big[\mathrm{E}\big(\alpha_1 * X_1 \cdot \alpha_2 * X_2 | \sigma(X_1, X_2)\big) \\
&\quad - \mathrm{E}\big(\alpha_1 * X_1 | \sigma(X_1, X_2)\big)\mathrm{E}\big(\alpha_2 * X_2 | \sigma(X_1, X_2)\big)\big] \\
&\quad + \mathrm{Cov}\big(\alpha_1 * X_1, \alpha_2 * X_2\big) \\
&= \mathrm{E}\left[\mathrm{E}\left(\sum_{i=1}^{X_1} N_i \sum_{j=1}^{X_2} M_j \,\middle|\, \sigma(X_1, X_2)\right) - \alpha_1 X_1 \cdot \alpha_2 X_2\right] \\
&\quad + \mathrm{E}\big[\alpha_1 X_1 \cdot \alpha_2 X_2\big] - \mathrm{E}\big[\alpha_1 X_1\big]\mathrm{E}\big[\alpha_2 X_2\big] \\
&= \mathrm{E}\left[\mathrm{E}\left(\sum_{i=1}^{X_1} N_i \,\middle|\, \sigma(X_1, X_2)\right) \mathrm{E}\left(\sum_{j=1}^{X_2} M_j \,\middle|\, \sigma(X_1, X_2)\right)\right] \\
&\quad - \mathrm{E}\big[\alpha_1 X_1\big]\mathrm{E}\big[\alpha_2 X_2\big] \\
&= \mathrm{E}\big[\mathrm{E}\left(\alpha_1 * X_1 \,|\, \sigma(X_1, X_2)\right) \mathrm{E}\left(\alpha_2 * X_2 \,|\, \sigma(X_1, X_2)\right)\big] \\
&\quad - \mathrm{E}\big[\alpha_1 X_1\big]\mathrm{E}\big[\alpha_2 X_2\big] \\
&= \mathrm{E}\big[\alpha_1 X_1 \cdot \alpha_2 X_2\big] - \mathrm{E}\big[\alpha_1 X_1\big]\mathrm{E}\big[\alpha_2 X_2\big] \\
&= \alpha_1 \alpha_2 \mathrm{Cov}(X_1, X_2),
\end{aligned}$$

where $\sigma(X_1, X_2)$ denotes the sigma algebra generated for X_1 and X_2, and for the hypothesis the counting series N_i's of $\alpha_1 * X_1$ are independent of counting series M_j's of $\alpha_2 * X_2$.

viii) $\mathrm{E}[(\alpha_1 * X_1)^2] = \mathrm{E}\{\mathrm{E}[(\alpha_1 * X_1)^2 \,|X_1]\} = \mathrm{E}[\alpha_1 X_1 + (\alpha_1 X_1)^2] = \alpha_1 \mathrm{E}[X_1] + \alpha_1^2 \mathrm{E}[X_1^2]$.

ix) From equation (4.4) we have

$$\begin{aligned} \mathrm{E}[(\alpha_1 * X_1)^r] &= \mathrm{E}\{\mathrm{E}[(\alpha_1 * X_1)^r \,|\, X_1]\} \\ &= \mathrm{E}\{\mathrm{E}[(\mathrm{Poisson}(\alpha_1 X_1))^r]\} \\ &= \sum_{k=1}^{r} S(r,k)\, \alpha_1^k \,\mathrm{E}[X_1^k]. \end{aligned}$$

x)

$$\begin{aligned} \mathrm{E}[(\alpha_1 * X_1) X_2] &= \mathrm{E}\{\mathrm{E}[(\alpha_1 * X_1) X_2 \,|\, \sigma(X_1, X_2)]\} \\ &= \mathrm{E}\{X_2 \,\mathrm{E}[(\alpha_1 * X_1) \,|\, \sigma(X_1, X_2)]\} \\ &= \mathrm{E}[X_2 \cdot \alpha_1 X_1] = \alpha_1 \mathrm{E}[X_1 X_2]. \end{aligned}$$

xi)

$$\begin{aligned} \mathrm{E}[(\alpha_1 * X_1)^2 X_2] &= \mathrm{E}\{\mathrm{E}[(\alpha_1 * X_1)^2 X_2 \,|\, \sigma(X_1, X_2)]\} \\ &= \mathrm{E}\{X_2 \,\mathrm{E}[(\alpha_1 * X_1)^2 \,|\, \sigma(X_1, X_2)]\} \\ &= \mathrm{E}\left[X_2 \left(\alpha_1 X_1 + \alpha_1^2 X_1^2\right)\right] \\ &= \alpha_1 \mathrm{E}[X_1 X_2] + \alpha_1^2 \mathrm{E}[X_1^2 X_2]. \end{aligned}$$

xii)

$$\begin{aligned} \mathrm{E}[(\alpha_1 * X_1)^r X_2] &= \mathrm{E}\{\mathrm{E}[(\alpha_1 * X_1)^r X_2 \,|\, \sigma(X_1, X_2)]\} \\ &= \mathrm{E}\{X_2 \,\mathrm{E}[(\alpha_1 * X_1)^r \,|\, \sigma(X_1, X_2)]\} \\ &= \mathrm{E}\left[X_2 \sum_{k=1}^{r} S(r,k)\, \alpha_1^k X_1^k\right] \\ &= \sum_{k=1}^{r} S(r,k)\, \alpha_1^k \,\mathrm{E}[X_1^k X_2]. \end{aligned}$$

xiii)

$$
\begin{aligned}
\mathrm{E}\left[(\alpha_1 * X_1)\prod_{i=2}^{m} X_i\right] &= \mathrm{E}\left\{\mathrm{E}\left[(\alpha_1 * X_1)\prod_{i=2}^{m} X_i \,\middle|\, \sigma(X_1, X_2, \ldots, X_m)\right]\right\} \\
&= \mathrm{E}\left\{\prod_{i=2}^{m} X_i \cdot \mathrm{E}\left[(\alpha_1 * X_1) \,|\, \sigma(X_1, X_2, \ldots, X_m)\right]\right\} \\
&= \mathrm{E}\left[\prod_{i=2}^{m} X_i \cdot \alpha_1 X_1\right] \\
&= \alpha_1 \mathrm{E}\left[\prod_{i=1}^{m} X_i\right],
\end{aligned}
$$

where $\sigma(X_1, X_2, \ldots, X_m)$ denotes the sigma algebra generated by $X_1, X_2, \ldots, X_m$.

xiv)

$$
\begin{aligned}
\mathrm{E}\left[(\alpha_1 * X_1)^r\prod_{i=2}^{m} X_i\right] &= \mathrm{E}\left\{\mathrm{E}\left[(\alpha_1 * X_1)^r\prod_{i=2}^{m} X_i \,\middle|\, \sigma(X_1, X_2, \ldots, X_m)\right]\right\} \\
&= \mathrm{E}\left\{\prod_{i=2}^{m} X_i \cdot \mathrm{E}\left[(\alpha_1 * X_1)^r \,|\, \sigma(X_1, X_2, \ldots, X_m)\right]\right\} \\
&= \mathrm{E}\left[\prod_{i=2}^{m} X_i \cdot \sum_{k=1}^{r} S(r,k)\, \alpha_1^k X_1^k\right] \\
&= \sum_{k=1}^{r} S(r,k)\, \alpha_1^k \mathrm{E}\left[X_1^k \prod_{i=1}^{m} X_i\right].
\end{aligned}
$$

xv)

$$
\begin{aligned}
\mathrm{E}[(\alpha_1 * X_1)(\alpha_2 * X_2)] &= \mathrm{E}\left\{\mathrm{E}\left[(\alpha_1 * X_1)(\alpha_2 * X_2) \,|\, \sigma(X_1, X_2)\right]\right\} \\
&= \mathrm{E}\left\{\mathrm{E}\left[\sum_{i=1}^{X_1} N_i \cdot \sum_{j=1}^{X_2} M_j \,\middle|\, \sigma(X_1, X_2)\right]\right\} \\
&= \mathrm{E}\left\{\mathrm{E}\left[\sum_{i=1}^{X_1} N_i \,\middle|\, \sigma(X_1, X_2)\right] \mathrm{E}\left[\sum_{j=1}^{X_2} M_j \,\middle|\, \sigma(X_1, X_2)\right]\right\} \\
&= \mathrm{E}\left\{\mathrm{E}\left[\alpha_1 * X_1 \,|\, \sigma(X_1, X_2)\right] \mathrm{E}\left[\alpha_2 * X_2 \,|\, \sigma(X_1, X_2)\right]\right\} \\
&= \alpha_1 \alpha_2 \mathrm{E}[X_1 X_2],
\end{aligned}
$$

where the N_i's and the M_j's are the counting series of $\alpha_1 * X_1$ and $\alpha_2 * X_2$ respectively, and they are independent for the hypothesis.

Property *xvi)* follow by a similar argument.

xvii)

$$\begin{aligned}
\mathrm{E}[(\alpha_1 * X_1)^2(\alpha_2 * X_2)] &= \mathrm{E}\left\{ \mathrm{E}\left[(\alpha_1 * X_1)^2(\alpha_2 * X_2) \,\middle|\, \sigma(X_1, X_2) \right] \right\} \\
&= \mathrm{E}\left\{ \mathrm{E}\left[\left(\sum_{i=1}^{X_1} N_i \right)^2 \cdot \sum_{j=1}^{X_2} M_j \,\middle|\, \sigma(X_1, X_2) \right] \right\} \\
&= \mathrm{E}\left\{ \mathrm{E}\left[\left(\sum_{i=1}^{X_1} N_i \right)^2 \,\middle|\, \sigma(X_1, X_2) \right] \mathrm{E}\left[\sum_{j=1}^{X_2} M_j \,\middle|\, \sigma(X_1, X_2) \right] \right\} \\
&= \mathrm{E}\left\{ \mathrm{E}\left[(\alpha_1 * X_1)^2 \,\middle|\, \sigma(X_1, X_2) \right] \mathrm{E}\left[\alpha_2 * X_2 \,\middle|\, \sigma(X_1, X_2) \right] \right\} \\
&= \mathrm{E}\left[(\alpha_1 X_1 + \alpha_1^2 X_1^2)(\alpha_2 X_2) \right] \\
&= \alpha_1 \alpha_2 \mathrm{E}[X_1 X_2] + \alpha_1^2 \alpha_2^2 \mathrm{E}[X_1^2 X_2].
\end{aligned}$$

Property *xviii)* follow by a similar argument.

xix)

$$\begin{aligned}
\mathrm{E}[\alpha_1 * (\alpha_1 * X_1)] &= \mathrm{E}\left\{ \mathrm{E}\left[\alpha_1 * (\alpha_1 * X_1) \,|\, X_1 \right] \right\} \\
&= \mathrm{E}\left\{ \mathrm{E}\left[\alpha_1 * \mathrm{Poisson}(\alpha_1 X_1) \right] \right\} \\
&= \mathrm{E}\left\{ \alpha_1 \mathrm{E}\left[\mathrm{Poisson}(\alpha_1 X_1) \right] \right\} \\
&= \alpha_1^2 \mathrm{E}[X_1].
\end{aligned}$$

xx) Proof by mathematical induction

- ⋆ For k=1 we have that $\mathrm{E}[\alpha_1 * X_1] = \alpha_1 \mathrm{E}[X_1]$ by property *iii)*
- ⋆ Suppose for $k = r - 1$ that

$$\mathrm{E}[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * X_1))}_{r-1\ \alpha_1\text{'s}}] = \alpha_1^{r-1} \mathrm{E}[X_1]$$

- ⋆ For $k = r$ we have

$$\mathrm{E}[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * X_1))}_{r\ \alpha_1\text{'s}}] = \mathrm{E}[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * Y))}_{r-1\ \alpha_1\text{'s}}]$$

$$
\begin{aligned}
&= \alpha_1^{r-1}\,\mathrm{E}[\,Y\,] \\
&= \alpha_1^{r-1}\,\mathrm{E}[\,\alpha_1 * X_1\,] \\
&= \alpha_1^{r}\,\mathrm{E}[\,X_1\,],
\end{aligned}
$$

where $Y = \alpha_1 * X_1$.

xxi) Proof by mathematical induction

- ⋆ For k=1 we have that $\mathrm{Var}[\alpha_1 * X_1] = \alpha_1^2\,\mathrm{Var}[X_1] + \alpha_1\,\mathrm{E}[X_1]$ by property *(iv.)*
- ⋆ Suppose for $k = r-1$ that

$$
\mathrm{Var}[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * X_1))}_{r-1\ \ \alpha_1\text{'s}}\,] = \alpha_1^{2(r-1)}\,\mathrm{Var}[\,X_1\,] + \left(\frac{1-\alpha_1^{r-1}}{1-\alpha_1}\right)\alpha_1^{r-1}\,\mathrm{E}\,[\,X_1\,]
$$

- ⋆ For $k = r$ we have

$$
\begin{aligned}
\mathrm{Var}[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * X_1))}_{r\ \ \alpha_1\text{'s}}\,] &= \mathrm{Var}[\underbrace{\alpha_1 * \cdots (\alpha_1 * (\alpha_1 * Y))}_{r-1\ \ \alpha_1\text{'s}}\,] \\
&= \alpha_1^{2(r-1)}\,\mathrm{Var}[\,Y\,] + \left(\frac{1-\alpha_1^{r-1}}{1-\alpha_1}\right)\alpha_1^{r-1}\,\mathrm{E}\,[\,Y\,] \\
&= \alpha_1^{2(r-1)}\left(\alpha_1^2\,\mathrm{Var}[X_1] + \alpha_1\,\mathrm{E}[X_1]\right) \\
&\quad + \left(\frac{1-\alpha_1^{r-1}}{1-\alpha_1}\right)\alpha_1^{r-1}\,\alpha_1\mathrm{E}\,[\,X_1\,] \\
&= \alpha_1^{2r}\,\mathrm{Var}[X_1] + \alpha_1^{2r-1}\,\mathrm{E}[X_1] + \left(\frac{1-\alpha_1^{r-1}}{1-\alpha_1}\right)\alpha_1^{r}\,\mathrm{E}\,[X_1] \\
&= \alpha_1^{2r}\,\mathrm{Var}[X_1] + \left(\frac{1-\alpha_1^{r}}{1-\alpha_1}\right)\alpha_1^{r}\,\mathrm{E}\,[X_1],
\end{aligned}
$$

where $Y = \alpha_1 * X_1$.

Properties *xxii)* and *xxiii)* follow from properties *xx)* and *xxi)* respectively.

References

Adell, J. A. and Jodrá, P. (2005). The median of the Poisson distribution. *Metrika*, 61(3):337–346.

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8(3):261–275.

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(1):19–30.

Brännäs, K. and Johansson, P. (1994). Time series count data regression. *Communications in Statistics - Theory and Methods*, 23(10):2907–2925.

Chen, J. and Rubin, H. (1986). Bounds for the difference between median and mean of Gamma and Poisson distributions. *Statistics & Probability Letters*, 4(6):281–283.

Choi, K. P. (1994). On the medians of Gamma distributions and an equation of Ramanujan. *Proceedings of the American Mathematical Society*, 121(1):245–251.

Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, 87(3):491–505.

Drost, F. C., Van Den Akker, R., and Werker, B. J. M. (2008). Local asymptotic normality and efficient estimation for INAR(p) models. *Journal of Time Series Analysis*, 29(5):783–801.

Du, J. and Li, Y. (1991). The integer-valued autoregressive INAR(p) model. *Journal of Time Series Analysis*, 12(2):129–142.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, 27(6):923–942.

Franke, J. and Seligmann, T. (1992). *Conditional Maximum-likelihood Estimates for INAR(1)-processes and Their Application to Modelling Epileptic Seizure Counts*. Preprint. Sonderforschungsbereich 123, Univ.

Freeland, R. K. (1998). *Statistical Analysis of Discrete Time Series with Applications to the Analysis of Workers Compensation Claims Data*. PhD Thesis, The University of British Columbia, Canada.

Freeland, R. K. and McCabe, B. P. M. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20(3):427–434.

Freeland, R. K. and McCabe, B. P. M. (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics & Probability Letters*, 73(2):147–153.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press, New York.

Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. CORE Discussion Paper, 2003-63, University of Louvain, Belgium.

Jorgensen, B., Lundbye-Christensen, S., Song, P. X., and Sun, L. (1999). A state space model for multivariate longitudinal count data. *Biometrika*, 86(1):169–181.

Jung, R. C., Ronning, G., and Tremayne, A. R. (2005). Estimation in conditional first order autoregression with discrete support. *Statistical Papers*, 46(2):195–224.

Jung, R. C. and Tremayne, A. R. (2006). Coherent forecasting in integer time series models. *International Journal of Forecasting*, 22(2):223–238.

Kim, H. Y. and Park, Y. (2008). A non-stationary integer-valued autoregressive model. *Statistical Papers*, 49(3):485–502.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6(3):629–642.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, 19(4):439–455.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, 21(4):645–650.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, 20(4):822.

McKenzie, E. (2003). Ch. 16. Discrete variate time series. In Shanbhag, D. and Rao, C., editors, *Stochastic Processes: Modelling and Simulation*, volume 21 of *Handbook of Statistics*, pages 573 – 606. Elsevier.

Pitman, J. (1997). Some probabilistic aspects of set partitions. *Amer. Math. Monthly*, 104:201–209.

Ross, S. M. (1983). *Stochastic Processes*. Wiley.

Silva, I. and Silva, M. E. (2006). Asymptotic distribution of the yule-walker estimator for INAR(p) processes. *Statistics & Probability Letters*, 76(15):1655–1663.

Silva, I. M. M. (2005). *Contributions to the analysis of discrete-valued time series*. PhD Thesis, University of Porto, Portugal.

Silva, M. E. and Oliveira, V. (2004). Difference Equations for the Higher Order Moments and Cumulants of the INAR(1) Model. *Journal of Time Series Analysis*, 25(3):317–333.

Silva, N., Pereira, I., and Silva, M. E. (2009). Forecasting in INAR(1) model. *RevStat Journal*, 7(3):119–134.

Steutel, F. W. and Harn, K. v. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7(5):893–899.

Tocher, K. D. (1963). *The Art of Simulation*. English University Press, London.

Weiß, C. (2008). Thinning operations for modeling time series of counts - a survey. *AStA Advances in Statistical Analysis*, 92(3):319–341.

Weiß, C. (2009). Modelling time series of counts with overdispersion. *Statistical Methods and Applications*, 18(4):507–519.

Weiß, C. (2010). The INARCH(1) model for overdispersed time series of counts. *Communications in Statistics - Simulation and Computation*, 39(6):1269–1291.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, 75(4):621–629.

Zelen, M. and Severo, N. C. (1964). Probability functions. *Handbook of Mathematical Functions, Edited by M. Abramowitz and I. A. Stegun, Applied Mathematics Series*, 55:925–995.

Zhu, F. and Wang, D. (2011). Estimation and testing for a Poisson autoregressive model. *Metrika*, 73:211–230.