



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

ISAAC JALES COSTA SOUZA

PROCESSO AUTOREGRESSIVO PARA VALORES INTEIROS COM
INOVAÇÕES INFLACIONADAS DE ZEROS

Recife

2020

ISAAC JALES COSTA SOUZA

PROCESSO AUTOREGRESSIVO PARA VALORES INTEIROS COM
INOVAÇÕES INFLACIONADAS DE ZEROS

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

Área de Concentração: Estatística Aplicada

Orientador: Prof. Dr. Aldo William Medina Garay

Coorientadora: Prof^a. Dr^a. Francielle de Lima Medina

Recife

2020

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

S729p Souza, Isaac Jales Costa
Processo autoregressivo para valores inteiros com inovações inflacionadas de zeros / Isaac Jales Costa Souza. – 2020.
112 f.: il., fig., tab.

Orientador: Aldo William Medina Garay.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2020.
Inclui referências e apêndices.

1. Estatística matemática. 2. Algoritmo EM. I. Garay, Aldo William Medina (orientador). II. Título.

519.5 CDD (23. ed.) UFPE- CCEN 2020 - 56

ISAAC JALES COSTA SOUZA

**PROCESSO AUTOREGRESSIVO PARA VALORES INTEIROS COM INOVAÇÕES
INFLACIONADAS DE ZEROS**

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

Aprovada em: 19 de Fevereiro de 2020.

BANCA EXAMINADORA

Prof.(º) Aldo William Medina Garay
UFPE

Prof.(º) Celso Rômulo Barbosa Cabral
UFAM

Prof.(º) Rodrigo Lambert
UFU

Prof.(º) Raydonal Ospina Martínez
UFPE

Prof.(º) Vinicius Quintas Souto Maior
UFPE

AGRADECIMENTOS

Aos meus pais, Paulo de Tarso Costa Souza e Altaíva Jales Costa Souza, por todo incentivo e apoio durante o período do Doutorado.

Ao meu orientador Aldo William Medina Garay e coorientadora Francielle de Lima Medina, por toda dedicação, ensinamentos e conselhos, os quais levarei para a toda minha vida acadêmica.

Aos professores André Pinho, Dione Valença, Paulo César Formiga, Bernardo Borba e Fidel Castro, os quais tiveram uma imensa contribuição para minha formação acadêmica ao longo da minha graduação e mestrado, na Universidade Federal do Rio Grande do Norte.

Aos meus chefes, durante o período de Doutorado, Enio Virgílio de Oliveira Matias e Rivaldo do Nascimento Júnior, por toda compreensão e apoio.

Aos meus amigos e colegas do PPGE: Jhonnata, Bruna, Cristiany, César, Cristina, Luis, João, Eduardo, Anabeth, Luan, Lucas, Marcelo, Pedro e Jodavid, pela amizade, apoio, companheirismo e por toda a ajuda proporcionada.

Ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco.

À CAPES pelo apoio financeiro.

RESUMO

Em algumas situações, séries temporais de contagem podem apresentar uma quantidade de zeros maior do que o esperado no modelo ajustado. Com o intuito de modelar tais séries, neste trabalho foi proposta uma classe de processos autorregressivos para valores inteiros não negativos ($\text{INAR}(p)$) com inovações seguindo distribuições inflacionadas de zeros, as quais denotamos por $\text{ZI-INAR}(p)$. Esta classe tem como caso particular o processo $\text{INAR}(p)$ com inovações Poisson Inflacionada de Zeros ($\text{ZIP-INAR}(p)$) e possibilita definir dois novos processos, com inovações Binomial Negativa Inflacionada de Zeros ($\text{ZINB-INAR}(p)$) e Poisson Inversa Gaussiana Inflacionada de Zeros ($\text{ZIPIG-INAR}(p)$). Para a classe, desenvolvemos a metodologia de estimação dos parâmetros via algoritmo EM e propomos métodos de bootstrap paramétrico e em blocos para construir intervalos de confiança e estimar a distribuição de previsão. Estudamos por meio de simulações o desempenho dos métodos de estimação propostos. Além disso, aplicações em conjunto de dados reais foram consideradas para comparar os processos $\text{ZIP-INAR}(p)$, $\text{ZINB-INAR}(p)$, e $\text{ZIPIG-INAR}(p)$, com os processos com inovações Poisson, Binomial Negativa e Poisson Inversa Gaussiana, via medidas de qualidade de ajuste e medidas de acurácia de previsão.

Palavras-chave: Algoritmo EM. Bootstrap em blocos. Distribuição inflacionada de zeros. Previsão. Poisson-Inversa Gaussiana. Processo autorregressivo de valores inteiros.

ABSTRACT

In some situations, counting time series may have more zero values than expected under the fitted model. In order to lead with such series, in this work, we propose a class of integer valued non-negative autoregressive processes of order p (INAR(p)) with innovations following zero inflated distributions, which we denoted by ZI-INAR(p). This class has as a particular case the INAR(p) process with Zero Inflated Poisson innovations (ZIP-INAR(p)) and it allows to define two new processes, with Zero Inflated Negative Binomial innovations (ZINB-INAR(p)) and Zero Inflated Gaussian Inverse innovations (ZIPIG-INAR(p)). For the class, we propose the parameters estimation methods via EM algorithm and the parametric and block bootstrap methods to obtain the confidence intervals and estimate the forecasting distribution. We studied through of simulations the performance of the parameters estimation. Also, we applied in a real data set our methodology in order to compare the performance of the ZIP-INAR(p), ZINB-INAR(p) and ZIPIG-INAR(p) processes, with the processes with Poisson, Negative Binomial and Poisson Inverse Gaussian innovations, via goodness-of-fit statistics and measures of forecasting accuracy.

Keywords: Block bootstrap. EM algorithm. Forecast. Integer valued autoregressive processes. Poisson Inverse Gaussian distribution. Zero inflated distribution.

LISTA DE FIGURAS

Figura 1 – Série <i>número de crimes no distrito 2206</i>	56
Figura 2 – Distribuição dos estimadores dos parâmetros do processo ZIPIG-INAR(1) obtidas pelos métodos de bootstrap paramétrico (B-PAR) e bootstrap circular em blocos (BCB).	58
Figura 3 – Distribuição de previsão $h = 1, 2$ e 3 passos à frente para a série temporal <i>número de crimes no distrito 2206</i>	60
Figura 4 – Série <i>número de crimes no distrito 1303</i>	86
Figura 5 – Distribuição dos estimadores dos parâmetros do processo ZIPIG-INAR(2) obtidas pelos métodos de bootstrap paramétrico (B-PAR) e bootstrap circular em blocos (BCB).	88
Figura 6 – Distribuição de previsão $h = 1, 2$ e 3 passos à frente para a série temporal <i>número de crimes no distrito 1303</i>	90
Figura 7 – Box-plot das estimativas dos parâmetros para os processos Po-INAR(1) e ZIP-INAR(1) considerando $\mu = 2$ e $n = 300$. Linha horizontal contínua preta: verdadeiro valor do parâmetro.	96
Figura 8 – Box-plot das estimativas dos parâmetros para os processos NB-INAR(1) e ZINB-INAR(1) considerando $\mu = 2$ e $n = 300$. Linha horizontal contínua preta: verdadeiro valor do parâmetro.	97
Figura 9 – Box-plot das estimativas dos parâmetros para os processos PIG-INAR(1) e ZIPIG-INAR(1) considerando $\mu = 2$ e $n = 300$. Linha horizontal contínua preta: verdadeiro valor do parâmetro.	98

LISTA DE TABELAS

Tabela 1	– Função geradora de probabilidade de distribuições ZI	20
Tabela 2	– VR e REQMR das estimativas de máxima verossimilhança do processo ZIP-INAR(1) para diferentes tamanhos amostrais e diferentes valores de π , fixando $\alpha = 0.3$ e $\lambda = 2$	51
Tabela 3	– VR e REQMR das estimativas de máxima verossimilhança do processo ZINB-INAR(1) para diferentes tamanhos amostrais e diferentes valores de π , fixando $\alpha = 0.3$, $\mu = 2$ e $\phi = 0.75, 1.5$ e 2.5	52
Tabela 4	– VR e REQMR das estimativas de máxima verossimilhança do processo ZIPIG-INAR(1) para diferentes tamanhos amostrais e diferentes valores de π , fixando $\alpha = 0.3$, $\mu = 2$ e $\phi = 0.75, 1.5$ e 2.5	52
Tabela 5	– VR das estimativas de máxima verossimilhança dos processos Po-INAR(1) e ZIP-INAR(1) em amostras de tamanho $n = 300$ para diferentes valores de π , fixando $\alpha = 0.3$ e $\lambda = 2$	54
Tabela 6	– VR das estimativas de máxima verossimilhança dos processos NB-INAR(1) e ZINB-INAR(1) em amostras de tamanho $n = 300$ para diferentes valores de π e ϕ , fixando $\alpha = 0.3$, $\mu = 2$	54
Tabela 7	– VR das estimativas de máxima verossimilhança dos processos PIG-INAR(1) e ZIPIG-INAR(1) em amostras de tamanho $n = 300$ para diferentes valores de π e ϕ , fixando $\alpha = 0.3$, $\mu = 2$	55
Tabela 8	– Estatísticas descritivas da série <i>número de crimes no distrito 2206</i> . . .	56
Tabela 9	– Estimativa dos parâmetros para os processos INAR(1) com inovações Poisson, ZIP, NB, ZINB, PIG e ZIPIG. Série: <i>número de crimes no distrito 2206</i>	57
Tabela 10	– Estimativa da probabilidade de zero e do valor esperado do <i>número de zeros consecutivos até o primeiro valor diferente de zero</i>	57
Tabela 11	– Estimativas e intervalos de confiança para os parâmetros do processo ZIPIG-INAR(1). B-PAR (Bootstrap paramétrico). BCB (Bootstrap circular em blocos).	59
Tabela 12	– Erro Absoluto Médio de Previsão 1 Passo à Frente (EAMP(1)) para a série <i>número de crimes no distrito 2206</i>	60

Tabela 13 – VR e REQMR das estimativas dos parâmetros do processo ZIP-INAR(p) para diferentes tamanhos de amostra, fixando $\pi = 0.3$ e 0.6 e $\lambda = 2$, considerando as ordens $p = 2$, com $\alpha = (0.3, 0.2)$, $p = 3$, com $\alpha = (0.25, 0.2, 0.15)$ e $p = 4$, com $\alpha = (0.25, 0.2, 0.15, 0.1)$	84
Tabela 14 – VR e REQMR das estimativas dos parâmetros do processo ZINB-INAR(p) para diferentes tamanhos de amostra, fixando $\pi = 0.3$ e 0.6 , $\mu = 2$, e $\phi = 1.5$, considerando as ordens $p = 2$, com $\alpha = (0.3, 0.2)$, $p = 3$, com $\alpha = (0.25, 0.2, 0.15)$ e $p = 4$, com $\alpha = (0.25, 0.2, 0.15, 0.1)$	84
Tabela 15 – VR e REQMR das estimativas dos parâmetros do processo ZIPIG-INAR(p) para diferentes tamanhos de amostra, fixando $\pi = 0.3$ e 0.6 , $\mu = 2$, e $\phi = 1.5$, considerando as ordens $p = 2$, com $\alpha = (0.3, 0.2)$, $p = 3$, com $\alpha = (0.25, 0.2, 0.15)$ e $p = 4$, com $\alpha = (0.25, 0.2, 0.15, 0.1)$	85
Tabela 16 – Estatísticas descritivas da série <i>número de crimes no distrito 1303</i> . . .	85
Tabela 17 – Estimativa dos parâmetros para os processos INAR(p) com inovações Poisson, ZIP, NB, ZINB, PIG e ZIPIG. Série: <i>crimes no distrito 1303</i> . . .	87
Tabela 18 – Estimativas e intervalos de confiança para os parâmetros do processo ZIPIG-INAR(2). B-PAR (Bootstrap paramétrico). BCB (Bootstrap circular em blocos).	87
Tabela 19 – EAMP(1) das $m = 12$ últimas observações da série <i>número de crimes no distrito 1303</i>	89

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVOS	13
2	CONCEITOS FUNDAMENTAIS	15
2.1	OPERADOR <i>THINNING</i> BINOMIAL	15
2.2	PROCESSO INAR(1)	16
2.3	DISTRIBUIÇÃO INFLACIONADA DE ZEROS	17
2.4	ALGORITMO EM	20
2.5	MÉTODOS BOOTSTRAP	22
3	PROCESSOS ZI-INAR(1)	24
3.1	DEFINIÇÃO E PROPRIEDADES	24
3.2	DISTRIBUIÇÃO MARGINAL E CONDICIONAL	29
3.3	DISTRIBUIÇÃO DE ZEROS	30
3.4	FUNÇÃO DE VEROSSIMILHANÇA	33
3.5	ESTIMAÇÃO DOS PARÂMETROS VIA ALGORITMO EM	34
3.6	INTERVALOS DE CONFIANÇA PARA OS PARÂMETROS	45
3.7	PREVISÃO	48
3.8	ESTUDOS DE SIMULAÇÃO	50
3.8.1	Simulação 1: Propriedades assintóticas	50
3.8.2	Simulação 2: Efeito da má especificação da distribuição das inovações	53
3.9	APLICAÇÃO	55
4	PROCESSOS ZI-INAR(p)	62
4.1	PROCESSO INAR(p)	62
4.2	DEFINIÇÃO E PROPRIEDADES	63
4.3	DISTRIBUIÇÃO CONDICIONAL	65
4.4	FUNÇÃO DE VEROSSIMILHANÇA	66
4.5	ESTIMAÇÃO DOS PARÂMETROS VIA ALGORITMO EM	67
4.6	PREVISÃO	82
4.7	ESTUDO DE SIMULAÇÃO	83
4.8	APLICAÇÃO	85
5	CONCLUSÕES E TRABALHOS FUTUROS	91

5.1	CONCLUSÕES	91
5.2	TRABALHOS FUTUROS	91
	REFERÊNCIAS	93
	APÊNDICE A – RESULTADOS DO ESTUDO DE SIMU- LAÇÃO 2 DO CAPÍTULO 3	96
	APÊNDICE B – DETALHES DO ALGORITMO EM DO PRO- CESSO ZIPIG-INAR	99
	APÊNDICE C – DEMONSTRAÇÃO DO TEOREMA 4.1 . .	105

1 INTRODUÇÃO

O interesse na modelagem de séries temporais de contagem tem crescido muito nos últimos anos devido sua ampla aplicabilidade em diferentes áreas do conhecimento. Por exemplo, podemos ter o interesse em estudar o número de acidentes de automóveis em uma rodovia, número de pacientes na fila de um hospital, número de indivíduos infectados por uma determinada doença em uma população, durante um determinado período de tempo, entre outros.

Um dos processos mais conhecidos para séries temporais de contagem é o processo autorregressivo de valores inteiros não negativos de primeira ordem, INAR(1), baseado no operador *thinning* Binomial, e introduzido por McKenzie (1985) e Al-Osh e Alzaid (1987), os quais consideraram que as inovações seguem distribuição Poisson. No entanto, na prática, séries temporais de contagem podem apresentar inflação de zeros, e neste caso, o processo INAR(1) pode não ser adequado. Considerando dados independentes, modelos com inflação de zeros têm sido utilizado em diversas áreas, como manufatura (LAMBERT, 1992), e agricultura (GARAY *et al.*, 2015). Com o intuito de modelar séries de contagem com inflação de zeros, Jazi, Jones e Lai (2012) propuseram o processo ZINAR(1), com inovações seguindo distribuição Poisson Inflacionada de Zeros. Barreto-Souza (2015) propôs o processo ZMGINAR(1), o qual supõe que o processo segue distribuição Geométrica Zero Modificada. Ristić, Bourguignon e Nastić (2019) propuseram o processo ZINGINAR(1), com distribuição marginal Geométrica Inflacionada de Zeros.

Apesar de terem sido propostos processos para séries temporais de contagem com inflação de zeros, estes processos não estão sob as mesmas condições. Isso porque o processo ZINAR(1) é baseado no operador *thinning* Binomial, enquanto que o processo ZMGINAR(1) utiliza o operador *thinning* Binomial Negativo e o processo ZINGINAR(1) é definido como o produto de um processo NGINAR(1) (RISTIĆ; BAKOUCH; NASTIĆ, 2009) com um processo de Bernoulli.

Para séries temporais de contagem com ordem maior que 1, Du e Li (1991) introduziram o processo autorregressivo de valores inteiros de ordem p , INAR(p). Nas últimas décadas, diversos estudos relacionados ao processo INAR(p) foram desenvolvidos, dentre estes podemos citar: Dion, Gauthier e Latour (1995), Latour (1998), Silva e Silva (2006), Bu e McCabe (2008), Bu, McCabe e Kaddour (2008). No entanto, não há muitos estudos para séries temporais com estrutura de dependência de ordem p e inflação de zeros.

Neste sentido, Maiti *et al.* (2014) propuseram o processo ZIPPAR(p), baseado no operador Pogram, o qual supõe que o processo segue distribuição Poisson Inflacionada de Zeros.

Nesta tese propomos uma classe de processos INAR(p), baseada no operador *thinning* Binomial, e com inovações seguindo uma distribuição Inflacionada de Zeros, a qual denotamos de ZI-INAR(p). Este trabalho generaliza o estudo de Jazi, Jones e Lai (2012), e possibilita propor que as inovações sigam distribuição Poisson Inflacionada de Zeros (ZIP), Binomial Negativa Inflacionada de Zeros (ZINB) e Poisson Inversa Gaussiana Inflacionada de Zeros (ZIPIG).

No Capítulo 2, definimos o processo INAR(1) e a classe de distribuições inflacionadas de zeros, e apresentamos um resumo do algoritmo EM e de métodos bootstrap. No Capítulo 3, definimos e apresentamos propriedades do processo ZI-INAR(1), propondo um algoritmo EM para obter os estimadores de máxima verossimilhança, métodos bootstrap para construir intervalos de confiança para os parâmetros e métodos de previsão. Além disso, apresentamos um estudo de simulação e uma aplicação a dados reais. No Capítulo 4, definimos o processo ZI-INAR(p), propondo também um algoritmo EM, intervalos de confiança bootstrap e metodologias de previsão, além de realizar um estudo de simulação e apresentar aplicações a dados reais. Por fim, conclusões e trabalhos futuros são discutidos no Capítulo 5.

1.1 OBJETIVOS

Esta tese tem como objetivo geral propor uma nova classe de processos autorregressivo de valores inteiros não negativos com inovações seguindo distribuições Inflacionadas de Zeros.

Considerando nosso objetivo geral do trabalho, apresentamos os seguintes objetivos específicos:

1. Propor um novo processo de primeira ordem, INAR(1), com inovações seguindo uma distribuição ZI (denotado por ZI-INAR(1)), incluindo um algoritmo EM para estimar os parâmetros por máxima verossimilhança, e intervalos de confiança bootstrap.
2. Desenvolver metodologias de previsão para o processo ZI-INAR(1) baseadas na distribuição condicional h passos à frente.
3. Propor um novo processo de ordem p , INAR(p), com inovações seguindo uma distribuição ZI (denotado por ZI-INAR(p)), generalizando os estudos de estimação e

previsão do processo ZI-INAR(1).

2 CONCEITOS FUNDAMENTAIS

Neste capítulo serão introduzidas algumas definições, notações e propriedades do processo autorregressivo de valores inteiros de primeira ordem (INAR(1)) e do modelo Inflacionado de Zeros, assim como algumas metodologias que serão úteis para o desenvolvimento do trabalho.

2.1 OPERADOR *THINNING* BINOMIAL

Definição 2.1. *Seja X uma variável aleatória (v.a.) não negativa de valores inteiros e $\alpha \in [0, 1]$. Então, o operador thinning Binomial “ $\circ$ ” é definido como:*

$$\alpha \circ X = \sum_{i=1}^X Z_i, \quad (2.1)$$

em que $\{Z_i\}_{i \in \mathbb{Z}}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (iid), seguindo uma distribuição Bernoulli, com $\mathbb{P}(Z_i = 1) = \alpha$, independente de X .

Diferentes tipos de operadores *thinning* têm sido propostos na literatura. Um exemplo é o operador *thinning* Binomial Negativo (RISTIĆ; BAKOUCH; NASTIĆ, 2009), que considera uma distribuição Geométrica para as v.a.’s Z_i ’s, definidas em (2.1). A utilização do operador *thinning* é frequente em processos estocásticos e em especial, o operador *thinning* Binomial é utilizado em processos de ramificação de Bienaymé–Galton–Watson (para mais detalhes, veja (FELLER, 2008, Capítulo 7)). Algumas propriedades do operador *thinning* Binomial são apresentadas a seguir:

Proposição 2.1. *Seja X uma v.a. não-negativa de valores inteiros e α e β valores reais pertencentes ao intervalo $[0, 1]$, então:*

- (i) $0 \circ X = 0$,
- (ii) $1 \circ X = X$,
- (iii) $(\alpha\beta) \circ X = \alpha \circ (\beta \circ X)$,
- (iv) $E(\alpha \circ X) = \alpha E(X)$,
- (v) $E((\alpha \circ X)^2) = \alpha^2 E(X^2) + \alpha(1 - \alpha)E(X)$,
- (vi) $Var(\alpha \circ X) = \alpha^2 Var(X) + \alpha(1 - \alpha)E(X)$.

Os itens (i)-(iii) seguem da Definição 2.1, enquanto que os itens (iv) e (v) foram demonstrados por Franke e Rao (1993), e (vi) é consequência imediata de (iv) e (v). Estas

e outras propriedades do operador *thinning* Binomial podem ser encontradas em Silva e Oliveira (2004).

2.2 PROCESSO INAR(1)

Considerando a Definição 2.1, o processo de contagem INAR(1) tem a seguinte representação estocástica:

$$Y_t = \alpha \circ Y_{t-1} + V_t, \quad t \in \mathbb{Z}, \quad (2.2)$$

em que $\{V_t\}_{t \in \mathbb{Z}}$ é uma sequência de variáveis aleatórias não negativas de valores inteiros independentes e identicamente distribuídas (iid), com média μ e variância finita σ^2 , definida como *inovações*, independentes de $Y_{t-1}, Y_{t-2}, \dots$ para todo $t \in \mathbb{Z}$.

É importante notar que, se $y_{t-1} > 0$, então $\alpha \circ Y_{t-1} | Y_{t-1} = y_{t-1} \sim \text{Bin}(y_{t-1}; \alpha)$, em que $\text{Bin}(a; b)$ denota a distribuição Binomial, com parâmetros ‘ a ’ e ‘ b ’, com $a \in \mathbb{N}^+$ e $b \in [0, 1]$, e se $y_{t-1} = 0$, então $\alpha \circ Y_{t-1} | Y_{t-1} = y_{t-1}$ segue uma distribuição degenerada em zero.

Segundo Al-Osh e Alzaid (1987), a v.a. “ $\alpha \circ Y_{t-1}$ ” pode ser interpretada como o número de elementos do instante $t-1$ que permaneceram no processo até o instante t , cada um com probabilidade α , enquanto que as inovações podem ser interpretadas como o número de elementos que foram introduzidos no processo, no intervalo de tempo $(t-1, t]$.

Conforme discutido por Al-Osh e Alzaid (1987) e Du e Li (1991), se $\alpha \in (0, 1)$, temos que $\{Y_t\}_{t \in \mathbb{Z}}$ é um processo estacionário, enquanto que $\alpha = 0$ implica em independência e $\alpha = 1$ implica em não estacionariedade. Além disso, se $0 \leq \alpha < 1$ o processo $\{Y_t\}_{t \in \mathbb{Z}}$ é estacionário de segunda ordem, com média e variância dadas por:

$$E(Y_t) = \frac{\mu}{1-\alpha} \text{ e } Var(Y_t) = \frac{\alpha\mu + \sigma^2}{1-\alpha^2}. \quad (2.3)$$

A função de autocovariância e autocorrelação de defasagem igual a h (ou *lag* h) são dadas por:

$$\gamma(h) = Cov(Y_t, Y_{t+h}) = \sigma^2 \alpha^h, \quad h = 0, \pm 1, \pm 2, \dots \quad (2.4)$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \alpha^h, \quad h = 0, \pm 1, \pm 2, \dots, \quad (2.5)$$

respectivamente. Note que a função de autocorrelação apresenta decaimento exponencial e é independente da distribuição da inovação do processo.

A distribuição marginal do processo INAR(1) pode ser representada em termos das inovações, utilizando substituições recursivas, que resultam em:

$$Y_t = \sum_{j=0}^{\infty} \alpha^j \circ V_{t-j}. \quad (2.6)$$

Conforme mencionado por Jazi, Jones e Lai (2012), em geral, é assumido que a inovação de um processo INAR(1) segue uma distribuição Poisson. No entanto, na prática, a incidência de zeros pode ser maior do que a esperada em uma distribuição Poisson. Um dos modelos mais conhecidos para dados de contagem inflacionado de zeros, é o modelo Poisson Inflacionado de Zeros (ZIP), em que o zero pode ter origem de duas fontes: a distribuição Poisson (zero amostral) ou a distribuição Bernoulli (zero estrutural). No contexto de séries temporais de valores inteiros, Jazi, Jones e Lai (2012) propuseram o processo INAR(1) com inovações seguindo distribuição ZIP.

Por outro lado, nos últimos anos, outros modelos inflacionados de zeros foram propostos na literatura, como por exemplo, a distribuição Binomial Negativa Inflacionada de Zeros (ZINB), a qual é obtida por uma mistura de uma distribuição Bernoulli com uma distribuição Binomial Negativa. Para mais detalhes, veja por exemplo, Hall (2000), Ridout, Hinde e Demétrio (2001), Garay *et al.* (2011).

Sendo assim, nosso objetivo no presente trabalho é introduzir uma classe de processos INAR, considerando inovações seguindo distribuições Inflacionada de Zeros (ZI), desenvolvendo algumas propriedades estruturais tais como média, variância e distribuição de probabilidade conjunta, além de propor um algoritmo tipo EM para estimar os parâmetros do processo, e métodos bootstrap para construir intervalos de confiança e realizar previsões.

2.3 DISTRIBUIÇÃO INFLACIONADA DE ZEROS

Definição 2.2. Dizemos que uma v.a. discreta V segue uma distribuição Inflacionada de Zeros (ZI), com parâmetros π e $\boldsymbol{\lambda}$, se tem a seguinte representação estocástica:

$$V = BU, \quad B \perp U,$$

em que B segue uma v.a. Bernoulli, com $\mathbb{P}(B = 1) = 1 - \mathbb{P}(B = 0) = 1 - \pi$, em que $0 \leq \pi < 1$ e U é uma v.a. discreta não negativa, com função de probabilidade (fp) dada por $h_U(u|\boldsymbol{\lambda})$, $\text{Var}(U) < \infty$. $\boldsymbol{\lambda}$ é um escalar ou vetor de parâmetros, indexado pela distribuição de U e $B \perp U$ indica que as v.a.'s B e U são independentes.

Como consequência imediata da Definição 2.2, podemos obter a fp da v.a. V , dada por:

$$\mathbb{P}(V = v) = \begin{cases} \pi + (1 - \pi)h_U(0|\boldsymbol{\lambda}), & v = 0; \\ (1 - \pi)h_U(v|\boldsymbol{\lambda}), & v = 1, 2, \dots, \end{cases} \quad (2.7)$$

em que $h_U(v|\boldsymbol{\lambda}) = \mathbb{P}(U = v)$.

Denotamos por $V \sim \text{ZI}(\pi, \boldsymbol{\lambda}; h_U)$ a v.a. que segue distribuição Inflacionada de Zeros. Neste caso, a média e a variância são obtidas por:

$$E(V) = (1 - \pi)E(U) \quad \text{e} \quad \text{Var}(V) = (1 - \pi)\left(\text{Var}(U) + \pi E^2(U)\right). \quad (2.8)$$

A distribuição da v.a. U determina a forma da distribuição ZI. A seguir descrevemos brevemente três casos particulares de modelos ZI:

- *Poisson Inflacionado de Zeros (ZIP):*

Neste caso, assume-se que a v.a. U segue uma distribuição Poisson, com parâmetro λ . Assim, a fp da v.a. V , definida na Equação (2.7), tem a forma:

$$\mathbb{P}(V = v) = \begin{cases} \pi + (1 - \pi)e^{-\lambda}, & v = 0; \\ (1 - \pi)\frac{e^{-\lambda}\lambda^v}{v!}, & v = 1, 2, \dots \end{cases}$$

Utilizamos a notação $V \sim \text{ZIP}(\pi, \lambda)$. Neste caso, por (2.8), temos:

$$E(V) = (1 - \pi)\lambda \quad \text{e} \quad \text{Var}(V) = \lambda(1 - \pi)(1 + \pi\lambda).$$

- *Binomial Negativo Inflacionado de Zeros (ZINB):*

Neste caso, U segue uma distribuição Binomial Negativa, denotada por $U \sim \text{NB}(\mu, \phi)$, com média $\mu > 0$, parâmetro de dispersão $\phi > 0$ e variância $\mu + \mu^2/\phi$. Sendo assim, $\boldsymbol{\lambda} = (\mu, \phi)$. Utilizamos a notação $V \sim \text{ZINB}(\pi, \mu, \phi)$ e a fp da v.a. V é dada por:

$$\mathbb{P}(V = v) = \begin{cases} \pi + (1 - \pi)\left(\frac{\phi}{\mu + \phi}\right)^\phi, & v = 0; \\ (1 - \pi)\frac{\Gamma(\phi + v)}{\Gamma(v + 1)\Gamma(\phi)}\left(\frac{\mu}{\mu + \phi}\right)^v\left(\frac{\phi}{\mu + \phi}\right)^\phi, & v = 1, 2, \dots, \end{cases}$$

em que $\Gamma(\cdot)$ representa a função gama. A média e a variância do modelo ZINB são:

$$E(V) = (1 - \pi)\mu \quad \text{e} \quad \text{Var}(V) = (1 - \pi)\mu\left(1 + \mu\phi^{-1} + \pi\mu\right),$$

respectivamente. Para mais detalhes, veja Ridout, Hinde e Demétrio (2001), Garay *et al.* (2011), Martinez-Flores, Bolfarine e Gomez (2013), Garay *et al.* (2015), entre outros.

- *Poisson Inversa Gaussiana Inflacionada de Zeros (ZIPIG)*

A distribuição ZIPIG é definida para o caso particular em que a v.a. U , dada na Equação (2.7), segue uma distribuição Poisson Inversa Gaussiana (PIG). Denotamos a v.a. com distribuição PIG por $U \sim \text{PIG}(\mu, \phi)$ com média $\mu > 0$ e variância $\mu + \mu^2/\phi$, sendo ϕ o parâmetro de dispersão. Neste caso, $\lambda = (\mu, \phi)$ e a fp da v.a. V , a qual denotamos por $V \sim \text{ZIPIG}(\pi, \mu, \phi)$, tem a seguinte forma:

$$\mathbb{P}(V = v) = \begin{cases} \pi + (1 - \pi)e^{\phi - \sqrt{\phi(\phi + 2\mu)}}, & v = 0; \\ (1 - \pi)\sqrt{\frac{2}{\pi}}[\phi(\phi + 2\mu)]^{-(v-1/2)/2} \\ \times \frac{e^{\phi}(\mu\phi)^v}{v!} K_{v-1/2}\left(\sqrt{\phi(\phi + 2\mu)}\right), & v = 1, 2, \dots \end{cases}$$

em que $K_\lambda(t) = \frac{1}{2} \int_0^\infty u^{\lambda-1} \exp\left\{-\frac{t}{2}\left(u + \frac{1}{u}\right)\right\} du$ é a função Bessel de terceira espécie (para mais detalhes, veja Abramowitz e Stegun (1964)). A média e a variância da distribuição ZIPIG(π, μ, ϕ) são dadas, respectivamente, por:

$$E(V) = (1 - \pi)\mu \quad \text{e} \quad \text{Var}(V) = (1 - \pi)\mu(1 + \mu\phi^{-1} + \pi\mu).$$

O modelo ZIPIG é um caso particular da distribuição de Mistura Poisson (MP) com $V|Z = z \sim \text{Poisson}(\mu z)$, em que Z segue uma distribuição Inversa Gaussiana com média 1 e parâmetro de dispersão ϕ , denotada por $Z \sim \text{IG}(1, \phi)$. Para mais detalhes, veja Karlis (2001) e Barreto-Souza e Simas (2016).

A seguir, definimos a função geradora de probabilidade (fgp) com o intuito de apresentar alguns resultados referentes às distribuições inflacionadas de zeros.

Definição 2.3. *Seja X uma v.a. inteira não negativa, então a fgp de X , denotada por $G_X(s)$, é definida como:*

$$G_X(s) = E(s^X), \quad 0 \leq s \leq 1.$$

A fp de X pode ser obtida por $\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}$, em que $G_X^{(k)}(s_0) = \frac{\partial^{(k)}}{\partial s^{(k)}} \{G_X(s)\} \Big|_{s=s_0}$. Conforme demonstrado por Johnson, Kemp e Kotz (2005), a fgp de uma v.a. $V \sim \text{ZI}(\pi, \lambda; h_U)$ satisfaz:

$$G_V(s) = \pi + (1 - \pi)G_U(s).$$

Para obter a fgp da v.a. “ $\alpha \circ V$ ”, utilizamos o fato que a fgp da soma $Z_1 + \dots + Z_X$, em que Z_i e X são as v.a.’s independentes, $\forall i \in \mathbb{Z}$, definidas em (2.1), é dada por (FELLER,

2008):

$$G_{Z_1+\dots+Z_X}(s) = G_X(G_{Z_1}(s)).$$

Assim, temos que:

$$G_{\alpha \circ X}(s) = G_X(1 - \alpha + \alpha s), \quad (2.9)$$

em que $G_{Z_1}(s) = 1 - \alpha + \alpha s$ é a fgp da v.a. Bernoulli. Portanto:

$$\begin{aligned} G_{\alpha \circ V}(s) &= G_V(1 - \alpha + \alpha s) \\ &= \pi + (1 - \pi)G_U(1 - \alpha + \alpha s) \\ &= \pi + (1 - \pi)G_{\alpha \circ U}(s), \end{aligned} \quad (2.10)$$

em que U é uma v.a. com função densidade $h_U(\cdot|\boldsymbol{\lambda})$. Segue de (2.10) que, se $V \sim \text{ZI}(\pi, \boldsymbol{\lambda}; h_U)$, então $\alpha \circ V \sim \text{ZI}(\pi, \boldsymbol{\lambda}^*; h_{\alpha \circ U})$, em que $\boldsymbol{\lambda}^* = (\alpha, \boldsymbol{\lambda})$, e $h_{\alpha \circ U}$ denota a fp da v.a. $\alpha \circ U$. A Tabela 1 apresenta a fgp da v.a. V seguindo uma distribuição Inflacionada de Zeros, considerada nessa seção, assim como a distribuição de $\alpha \circ V$ em cada caso. O resultado da terceira coluna pode ser facilmente demonstrado utilizando a relação (2.9).

Tabela 1 – Função geradora de probabilidade de distribuições ZI

Modelo	$G_V(s)$	Dist. de $\alpha \circ V$	$G_{\alpha \circ V}(s)$
$\text{ZIP}(\pi, \lambda)$	$\pi + (1 - \pi)e^{-\lambda(1-s)}$	$\text{ZIP}(\pi, \alpha\lambda)$	$\pi + (1 - \pi)e^{-\alpha\lambda(1-s)}$
$\text{ZINB}(\pi, \mu, \phi)$	$\pi + (1 - \pi) \left(\frac{\phi}{\phi + \mu(1-s)} \right)^\phi$	$\text{ZINB}(\pi, \alpha\mu, \phi)$	$\pi + (1 - \pi) \left(\frac{\phi}{\phi + \alpha\mu(1-s)} \right)^\phi$
$\text{ZIPIG}(\pi, \mu, \phi)$	$\pi + (1 - \pi) \times \exp \left\{ \phi \left[1 - \left(1 + \frac{2\mu}{\phi}(1-s) \right)^{1/2} \right] \right\}$	$\text{ZIPIG}(\pi, \alpha\mu, \phi)$	$\pi + (1 - \pi) \times \exp \left\{ \phi \left[1 - \left(1 + \frac{2\alpha\mu}{\phi}(1-s) \right)^{1/2} \right] \right\}$

2.4 ALGORITMO EM

O algoritmo EM proposto por Dempster, Laird e Rubin (1977) é uma ferramenta computacional utilizada para obter as estimativas de máxima verossimilhança por meio de um processo iterativo. A ideia do algoritmo EM é basicamente considerar a presença de “dados não observáveis” (ou variáveis latentes) com o intuito de simplificar o cálculo das estimativas dos parâmetros. Assim, sendo $\mathbf{y}$ o vetor dos “dados observáveis” e $\mathbf{s}$ o vetor das variáveis latentes, no algoritmo EM é obtida a expressão da função de log-verossimilhança

dos “dados aumentados” $\mathbf{y}_c = (\mathbf{y}, \mathbf{s})$, chamada de *função de log-verossimilhança completa* e denotada por $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$, em que $\boldsymbol{\theta}$ é o vetor de parâmetros. Na prática, não é possível obter a função de log-verossimilhança completa uma vez que não conhecemos as variáveis latentes. Sendo assim, obtemos o valor esperado de $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ e maximizamos este valor esperado com relação ao vetor de parâmetros $\boldsymbol{\theta}$, ou seja, o algoritmo EM é dado pelas seguintes etapas:

1. Passo E (Esperança): Calcule $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = E(\ell_c(\boldsymbol{\theta})|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)})$, em que $\hat{\boldsymbol{\theta}}^{(k)}$ é a estimativa de $\boldsymbol{\theta}$ na k -ésima iteração do algoritmo.
2. Passo M (Maximização): Calcule $\hat{\boldsymbol{\theta}}^{(k+1)} = \arg\max_{\boldsymbol{\theta} \in \Theta} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$, em que Θ é o espaço paramétrico do vetor $\boldsymbol{\theta}$.

Em algumas situações, a maximização simultânea das componentes do vetor $\boldsymbol{\theta}$ no passo M do algoritmo EM é um procedimento computacional complexo. Uma alternativa para contornar este problema é utilizar o algoritmo ECM proposto por Meng e Rubin (1993), que substitui o passo M do algoritmo EM por uma sequência de passos de maximização condicional (passo CM). Uma extensão do algoritmo ECM é o algoritmo ECME (LIU; RUBIN, 1994), que substitui os passos de maximização condicional por maximizações da função de verossimilhança marginal, isto é, da função de verossimilhança dos “dados incompletos”.

Conforme mencionado por Dempster, Laird e Rubin (1977), a função de verossimilhança $\ell(\boldsymbol{\theta}|\mathbf{Y})$ avaliada no ponto $\hat{\boldsymbol{\theta}}^{(k)}$ é não decrescente a cada iteração do algoritmo EM, e a sequência $\hat{\boldsymbol{\theta}}^{(k)}$ converge para um valor estacionário da função de verossimilhança observada sob certas condições de regularidade (para mais detalhes, ver Wu (1983) e Vaida (2005)).

Para inicializar o algoritmo, devemos estabelecer um valor inicial $\hat{\boldsymbol{\theta}}^{(0)}$ arbitrário para a estimativa de $\boldsymbol{\theta}$ na primeira iteração do algoritmo, e seguir os passos mencionados anteriormente de modo iterativo até que algum critério de convergência seja satisfeito. Por exemplo, esse critério de convergência pode ser estabelecido quando diferenças sucessivas de $\|\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}\|$ (ou de uma função do logaritmo da verossimilhança) for suficientemente pequena.

2.5 MÉTODOS BOOTSTRAP

O método bootstrap, introduzido por Efron (1979), é uma ferramenta estatística útil para realizar inferências sobre um parâmetro de interesse, tais como construir intervalos de confiança e testar hipóteses. O método consiste em obter amostras por reamostragem dos dados observados (bootstrap não-paramétrico) ou gerando valores de um modelo ajustado (bootstrap paramétrico). A partir destas amostras é possível estimar a distribuição de uma estatística de interesse τ .

Diversos métodos bootstrap foram propostos na literatura e em particular, para séries temporais, um dos mais utilizados é o bootstrap em blocos. A versão mais simples deste método, consiste em particionar a série em k blocos e posteriormente aleatorizar estes blocos até que se forme uma nova série temporal do tamanho da série observada. Assim, sendo $\mathbf{y} = (y_1, \dots, y_n)$ uma realização de uma série temporal com função densidade $f(\mathbf{y}|\boldsymbol{\theta})$ e $\hat{\boldsymbol{\theta}} = T(\mathbf{y})$ um estimador para $\boldsymbol{\theta}$, então o bootstrap em blocos é dado pelas seguintes etapas:

1. Escolha um tamanho l para os blocos, de tal modo que n/l seja um número inteiro, e defina os blocos $B_1, B_2, \dots, B_k$ por:

$$B_1 = (y_1, \dots, y_l), B_2 = (y_{l+1}, \dots, y_{2l}), \dots, B_k = (y_{n-l+1}, \dots, y_n),$$

em que $k = n/l$.

2. Selecione com reposição $k = n/l$ blocos para gerar a série temporal $\mathbf{y}^* = (y_1^*, \dots, y_{kl}^*)^\top$, em que $y_{1+(i-1)l}^*, y_{2+(i-1)l}^*, \dots, y_{il}^*$ correspondem as observações do i -ésimo bloco selecionado.
3. Calcule para a série temporal $\mathbf{y}^*$ a estimativa bootstrap $\hat{\boldsymbol{\theta}}^*$.
4. Repita os passos (1)-(3) R vezes, obtendo as estimativas $\{\hat{\boldsymbol{\theta}}_r^*, r = 1, \dots, R\}$.
5. Utilize a distribuição empírica das estimativas $\hat{\boldsymbol{\theta}}_1^*, \dots, \hat{\boldsymbol{\theta}}_R^*$ como a estimativa para a distribuição de $\hat{\boldsymbol{\theta}}$.

Note que os blocos permitem que a estrutura de dependência até a ordem l seja preservada. Outra versão do bootstrap em blocos, proposta por Kunsch (1989) e denotada por *bootstrap em blocos móveis*, considera blocos sobrepostos, ou seja, o primeiro bloco se inicia pela observação y_1 , o segundo, por y_2 , e assim por diante. Diversas variações do bootstrap em blocos para séries temporais foram propostas na literatura (para mais detalhes, veja Davison e Hinkley (1997)).

Sendo assim, no capítulo seguinte apresentamos o nosso processo INAR(1) com inovações seguindo distribuições ZI, em que propomos um algoritmo EM para estimar os parâmetros, além de métodos bootstrap para séries temporais, com o intuito de construir intervalos de confiança e realizar previsões.

3 PROCESSOS ZI-INAR(1)

Em processos INAR(1), geralmente assume-se que as inovações seguem distribuição Poisson. No entanto, conforme mencionamos anteriormente, séries temporais de contagem podem apresentar inflação de zeros. Diante disso, Jazi, Jones e Lai (2012) propuseram o processo INAR(1) com inovações seguindo distribuição ZIP. A proposta deste capítulo é desenvolver uma classe de processos autorregressivos para valores inteiros não negativos com inovações seguindo uma classe de distribuições Inflacionadas de Zeros, a qual denotamos por ZI-INAR(1), generalizando o trabalho por Jazi, Jones e Lai (2012). Conforme comentamos na introdução deste trabalho, a motivação em propor esta classe é que processos ZI-INAR(1) sempre estarão sob as mesmas condições, i.e., sempre serão baseados no mesmo operador (operador *thinning* Binomial), diferindo apenas quanto a distribuição da inovação adotada.

Este capítulo está organizado da seguinte maneira. Inicialmente, desenvolvemos algumas propriedades do processo ZI-INAR(1), assim como metodologias para estimação dos parâmetros e previsão. Em seguida, apresentamos dois estudos de simulação e finalizamos o capítulo com uma aplicação a dados reais.

3.1 DEFINIÇÃO E PROPRIEDADES

Definição 3.1. Dizemos que o processo $\{Y_t\}_{t \in \mathbb{Z}}$ é um processo autorregressivo para valores inteiros de primeira ordem, com inflação de zeros, denotado por ZI-INAR(1), se as inovações V_t seguem distribuição $ZI(\pi, \boldsymbol{\lambda}; h_{U_t})$, em que U_t é uma v.a. discreta e não negativa, h_{U_t} é a fp da v.a U_t , e $\boldsymbol{\lambda}$ é um vetor de parâmetros.

A distribuição inflacionada de zeros V_t determina o processo ZI-INAR(1). Sendo assim, denotamos de ZIP-INAR(1), ZINB-INAR(1) e ZIPIG-INAR(1) os processos ZI-INAR(1) com inovações ZIP, ZINB e ZIPIG, respectivamente. Por outro lado, denotamos de Po-INAR(1), os processos com inovações Poisson, NB-INAR(1), os processos com inovações Binomial Negativa, e PIG-INAR(1), os processos com inovações PIG. A seguir, apresentamos as expressões da esperança e da variância do processo ZI-INAR(1).

Proposição 3.1. Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo ZI-INAR(1) estacionário. Então, a esperança e a variância de $\{Y_t\}_{t \in \mathbb{Z}}$ são dadas por:

$$E(Y_t) = \frac{(1 - \pi)E(U_t)}{1 - \alpha} \quad (3.1)$$

e

$$Var(Y_t) = \frac{(1 - \pi) (\alpha E(U_t) + \pi E^2(U_t) + Var(U_t))}{1 - \alpha^2}, \quad (3.2)$$

em que U_t denota a v.a. com fp $h_{U_t}(\cdot|\lambda)$, para todo $t \in \mathbb{Z}$.

Prova. Utilizando a expressão da Esperança de uma v.a. inflacionada de zeros, temos que:

$$\begin{aligned} E(Y_t|Y_{t-1} = y_{t-1}) &= E(\alpha \circ Y_{t-1} + V_t|Y_{t-1} = y_{t-1}) \\ &= E(\alpha \circ Y_{t-1}|Y_{t-1} = y_{t-1}) + E(V_t|Y_{t-1} = y_{t-1}) \\ &= \alpha y_{t-1} + (1 - \pi) E(U_t). \end{aligned} \quad (3.3)$$

Assim,

$$\begin{aligned} E(Y_t) &= E(E(Y_t|Y_{t-1})) \\ &= E(\alpha Y_{t-1} + (1 - \pi) E(U_t)) \\ &= \alpha E(Y_{t-1}) + (1 - \pi) E(U_t). \end{aligned}$$

Utilizando o fato que o processo é estacionário, o que implica que a esperança de Y_t é constante para todo $t \in \mathbb{Z}$, obtemos:

$$E(Y_t) = \frac{(1 - \pi) E(U_t)}{1 - \alpha}.$$

Para obter a variância do processo, temos:

$$\begin{aligned} Var(Y_t|Y_{t-1} = y_{t-1}) &= Var(\alpha \circ Y_{t-1} + V_t|Y_{t-1} = y_{t-1}) \\ &= Var(\alpha \circ Y_{t-1}|Y_{t-1} = y_{t-1}) + Var(V_t|Y_{t-1} = y_{t-1}) \\ &= \alpha(1 - \alpha)y_{t-1} + (1 - \pi) (Var(U_t) + \pi E^2(U_t)). \end{aligned} \quad (3.4)$$

Assim, por (3.3) e (3.4), obtemos:

$$\begin{aligned} Var(Y_t) &= Var(E(Y_t|Y_{t-1})) + E(Var(Y_t|Y_{t-1})) \\ &= \alpha^2 Var(Y_{t-1}) + \alpha(1 - \alpha)E(Y_{t-1}) + (1 - \pi)(Var(U_t) + \pi E^2(U_t)) \\ &= \alpha^2 Var(Y_t) + \alpha(1 - \pi)E(U_t) + (1 - \pi)(Var(U_t) + \pi E^2(U_t)), \end{aligned} \quad (3.5)$$

em que a expressão (3.5) é obtida considerando que o processo é estacionário, e pela expressão da esperança dada em (3.1). Assim:

$$Var(Y_t) = \frac{(1 - \pi) (\alpha E(U_t) + \pi E^2(U_t) + Var(U_t))}{1 - \alpha^2}.$$

□

É possível verificar pelo resultado da Proposição 3.1 que a razão da variância pela média do processo $\{Y_t\}_{t \in \mathbb{Z}}$ pode ser expressa por:

$$\frac{Var(Y_t)}{E(Y_t)} = 1 + \frac{Var(U_t) - E(U_t) + \pi E^2(U_t)}{(1 + \alpha)E(U_t)}.$$

Note que, se $Var(U_t) \geq E(U_t)$, então $Var(Y_t)/E(Y_t) > 1$. Em particular, temos que:

- $V_t \sim \text{ZIP}(\pi, \lambda)$, então $U_t \sim \text{Poisson}(\lambda)$. Portanto:

$$E(Y_t) = \frac{(1 - \pi)\lambda}{1 - \alpha} \quad \text{e} \quad Var(Y_t) = \frac{(1 - \pi)\lambda[1 + \alpha + \pi\lambda]}{1 - \alpha^2},$$

e consequentemente:

$$\frac{Var(Y_t)}{E(Y_t)} = 1 + \frac{\pi\lambda}{1 + \alpha}.$$

- $V_t \sim \text{ZINB}(\pi, \mu, \phi)$ ou $V_t \sim \text{ZIPIG}(\pi, \mu, \phi)$, então a v.a. U_t segue uma distribuição $\text{NB}(\mu, \phi)$ ou $\text{PIG}(\mu, \phi)$, o que implica nos dois casos que $E(U_t) = \mu$ e $Var(U_t) = \mu + \mu^2/\phi$. Portanto:

$$E(Y_t) = \frac{(1 - \pi)\mu}{1 - \alpha} \quad \text{e} \quad Var(Y_t) = \frac{(1 - \pi)\mu[1 + \alpha + \mu(\pi + \phi^{-1})]}{1 - \alpha^2},$$

e consequentemente:

$$\frac{Var(Y_t)}{E(Y_t)} = 1 + \frac{\mu(\pi + \phi^{-1})}{1 + \alpha}.$$

A seguir, apresentamos a expressão da esperança condicional de Y_{t+h} dado $Y_t = y_t$, para $h = 1, 2, \dots$

Proposição 3.2. *Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo ZI-INAR(1) estacionário. Então, a esperança condicional de Y_{t+h} dado $Y_t = y_t$, $h = 1, 2, \dots$, é dada por:*

$$E(Y_{t+h}|Y_t = y_t) = \alpha^h y_t + (1 - \pi)E(U_t) \frac{1 - \alpha^h}{1 - \alpha}, \quad (3.6)$$

em que U_t denota a v.a. com fp $h_{U_t}(\cdot|\lambda)$, para todo $t \in \mathbb{Z}$.

Prova. *Utilizando substituições recursivas, Al-Osh e Alzaid (1987) representaram o processo INAR(1), de maneira geral, como:*

$$Y_{t+h} = \alpha^h \circ Y_t + \sum_{i=0}^{h-1} \alpha^i \circ V_{t+h-i}.$$

Assim, para processos $INAR(1)$ cujas inovações seguem distribuições ZI , obtemos:

$$\begin{aligned}
E(Y_{t+h}|Y_t = y_t) &= E\left(\alpha^h \circ Y_t + \sum_{i=0}^{h-1} \alpha^i \circ V_{t+h-i} | Y_t = y_t\right) \\
&= E\left(\alpha^h \circ Y_t | Y_t = y_t\right) + \sum_{i=0}^{h-1} E\left(\alpha^i \circ V_{t+h-i} | Y_t = y_t\right) \\
&= \alpha^h y_t + E(V_t) \sum_{i=0}^{h-1} \alpha^i \\
&= \alpha^h y_t + (1 - \pi) E(U_t) \frac{1 - \alpha^h}{1 - \alpha},
\end{aligned} \tag{3.7}$$

em que (3.7) é consequência do fato que as inovações V_{t+h-i} são independentes de Y_t , para $i = 0, 1, \dots, h-1$, e têm média constante, condicionado a $Y_t = y_t$. $\square$

A proposição a seguir afirma que os processos propostos para a nossa classe são autodecompostos (para detalhes relacionado a distribuição autodecomposta, veja Steutel e Harn (1979)). Esta condição garante a estacionariedade do processo. Como demonstrado por Alzaid e Al-Osh (1988), sob condições de estacionariedade, a fgp de um processo $INAR(1)$ pode ser expressa como:

$$G_{Y_t}(s) = G_{Y_t}(1 - \alpha + \alpha s) G_{V_t}(s). \tag{3.8}$$

A relação (3.8) satisfaz a definição de distribuição autodecomposta.

Proposição 3.3. *Os processos $INAR(1)$ com inovações pertencentes a classe das distribuições ZI , tais como ZIP , $ZINB$ e $ZIPIG$ são autodecompostos.*

Prova. *Para o processo $INAR(1)$ estacionário, temos que:*

$$Y_t = \sum_{j=0}^{\infty} \alpha^j \circ V_{t-j} = \sum_{j=1}^{\infty} \alpha^j \circ V_{t-j} + V_t.$$

Assim, utilizando o fato que as inovações $\{V_t\}_{t \in \mathbb{Z}}$ são independentes, temos que a fgp de Y_t pode ser expressa como:

$$G_{Y_t}(s) = \left[\prod_{j=1}^{\infty} G_{\alpha^j \circ V_{t-j}}(s) \right] G_{V_t}(s), \tag{3.9}$$

em que $G_{\alpha^j \circ V_{t-j}}(\cdot)$ e $G_{V_t}(\cdot)$ representam as fgp's de $\alpha^j \circ V_{t-j}$ e V_t , respectivamente, para $j \in \mathbb{N}^+$ e $t \in \mathbb{Z}$. Em particular,

- Se $V_t \sim ZIP(\pi, \lambda)$, o processo $ZIP-INAR(1)$ é autodecomposto (ver Jazi, Jones e Lai (2012) para mais detalhes).

- Se $V_t \sim \text{ZINB}(\pi, \mu, \phi)$, pela Tabela 2.1 (veja pág. 20), obtemos:

$$\begin{aligned}
\prod_{j=1}^{\infty} G_{\alpha^j \circ V_{t-j}}(s) &= \prod_{j=1}^{\infty} \left[\pi + (1 - \pi) \left(\frac{\phi}{\phi + \alpha^j \mu (1 - s)} \right)^{\phi} \right] \\
&= \prod_{j=1}^{\infty} \left[\pi + (1 - \pi) \left(\frac{\phi}{\phi + \mu (1 - (1 - \alpha^j + \alpha^j s))} \right)^{\phi} \right] \\
&= \prod_{j=1}^{\infty} G_{V_t}(1 - \alpha^j + \alpha^j s).
\end{aligned}$$

- Se $V_t \sim \text{ZIPIG}(\pi, \mu, \phi)$, pela Tabela 2.1 (veja pág. 20), obtemos:

$$\begin{aligned}
\prod_{j=1}^{\infty} G_{\alpha^j \circ V_{t-j}}(s) &= \prod_{j=1}^{\infty} \left[\pi + (1 - \pi) \exp \left\{ \phi \left[1 - \left(1 + \frac{2\alpha^j \mu}{\phi} (1 - s) \right)^{1/2} \right] \right\} \right] \\
&= \prod_{j=1}^{\infty} \left[\pi + (1 - \pi) \exp \left\{ \phi \left[1 - \left(1 + \frac{2\mu}{\phi} (1 - (1 - \alpha^j + \alpha^j s)) \right)^{1/2} \right] \right\} \right] \\
&= \prod_{j=1}^{\infty} G_{V_t}(1 - \alpha^j + \alpha^j s).
\end{aligned}$$

Assim, para os casos $\text{ZINB-INAR}(1)$ e $\text{ZIPIG-INAR}(1)$, obtemos de (3.9) que:

$$\begin{aligned}
G_{Y_t}(s) &= \left[\prod_{j=1}^{\infty} G_{\alpha^j \circ V_{t-j}}(s) \right] G_{V_t}(s) \\
&= \left[\prod_{j=1}^{\infty} G_{V_t}(1 - \alpha^j + \alpha^j s) \right] G_{V_t}(s).
\end{aligned} \tag{3.10}$$

Portanto, denotando $s^* = 1 - \alpha + \alpha s$, temos:

$$\begin{aligned}
G_{Y_t}(s) &= \left[\prod_{j=1}^{\infty} G_{V_t}(1 - \alpha^j + \alpha^j s) \right] G_{V_t}(s) \\
&= \left[\prod_{j=0}^{\infty} G_{V_t}(1 - \alpha^{j+1} + \alpha^{j+1} s) \right] G_{V_t}(s) \\
&= \left\{ \left[\prod_{j=1}^{\infty} G_{V_t}(1 - \alpha^j + \alpha^j (1 - \alpha + \alpha s)) \right] G_{V_t}(1 - \alpha + \alpha s) \right\} G_{V_t}(s) \\
&= \left\{ \left[\prod_{j=1}^{\infty} G_{V_t}(1 - \alpha^j + \alpha^j s^*) \right] G_{V_t}(s^*) \right\} G_{V_t}(s) \\
&= G_{Y_t}(s^*) G_{V_t}(s)
\end{aligned} \tag{3.11}$$

em que utilizamos (3.10) para obter (3.11), cuja expressão satisfaz a definição de distribuição autodecomposta dada por (3.8). Logo, os processos $\text{ZINB-INAR}(1)$ e $\text{ZIPIG-INAR}(1)$ são autodecompostos. $\square$

Note que o processo ZI-INAR(1) pode ser representado como a soma infinita das v.a.'s $\alpha^j \circ V_{t-j} \sim \text{ZI}(\pi, \boldsymbol{\lambda}_j^*)$, para todo $t \in \mathbb{Z}$, em que $\boldsymbol{\lambda}_j^* = (\alpha^j, \boldsymbol{\lambda}_j)$ (veja Eq. (2.6) e (2.10) – Cap. 2), com decaimento geométrico da média, isto é, $E(\alpha^{j+1} \circ V_{t-(j+1)}) / E(\alpha^j \circ V_{t-j}) = \alpha$ (veja Prop. 2.1 – Cap. 2), para todo $j \in \mathbb{Z}$. Este resultado é uma extensão de Jazi, Jones e Lai (2012), que demonstrou que a distribuição marginal de um processo INAR(1) com inovações ZIP pode ser expressa como a soma infinita de v.a.'s ZIP. Em particular, pelos resultados da Tabela 1 dada no Capítulo 2, a distribuição marginal de um processo ZINB-INAR(1) e ZIPIG-INAR(1) podem ser representadas pela soma infinita de v.a.'s ZINB e ZIPIG, respectivamente.

3.2 DISTRIBUIÇÃO MARGINAL E CONDICIONAL

A probabilidade condicional, $\mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1})$, do processo INAR(1) com inovações seguindo distribuições $\text{ZI}(\pi, \boldsymbol{\lambda}, h_{U_t})$, é dada por:

$$\begin{aligned}
 \mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}) &= \mathbb{P}(\alpha \circ Y_{t-1} + V_t = y_t | Y_{t-1} = y_{t-1}) \\
 &= \sum_{k=0}^{\min\{y_t, y_{t-1}\}} \mathbb{P}(\alpha \circ Y_{t-1} + V_t = y_t, \alpha \circ Y_{t-1} = k | Y_{t-1} = y_{t-1}) \\
 &= \sum_{k=0}^{\min\{y_t, y_{t-1}\}} \mathbb{P}(\alpha \circ Y_{t-1} = k | Y_{t-1} = y_{t-1}) \\
 &\quad \times \mathbb{P}(\alpha \circ Y_{t-1} + V_t = y_t | \alpha \circ Y_{t-1} = k, Y_{t-1} = y_{t-1}) \\
 &= \sum_{k=0}^{\min\{y_t, y_{t-1}\}} \mathbb{P}(\alpha \circ Y_{t-1} = k | Y_{t-1} = y_{t-1}) \times \mathbb{P}(V_t = y_t - k) \\
 &= \sum_{k=0}^{\min\{y_t, y_{t-1}\}} \binom{y_{t-1}}{k} \alpha^k (1 - \alpha)^{y_{t-1} - k} \\
 &\quad \times \left[\pi \mathbb{I}_{\{y_t - k = 0\}} + (1 - \pi) h_{U_t}(y_t - k | \boldsymbol{\lambda}) \right], \tag{3.12}
 \end{aligned}$$

em que $\mathbb{I}_{\{A\}}$ denota a função indicadora, isto é, $\mathbb{I}_{\{A\}} = 1$, se o evento A ocorre, e $\mathbb{I}_{\{A\}} = 0$, caso contrário. Note que o processo INAR(1) com inovações ZI é um processo de Markov, uma vez que o estado atual, y_t , depende apenas do estado anterior, y_{t-1} (veja Eq. 2.2 – Cap. 2). Sendo assim, a expressão (3.12) representa a probabilidade de transição.

A esperança e a variância condicional do processo foram obtidas em (3.3) e (3.4), e a distribuição marginal pode ser expressa em termos da probabilidade de transição

da seguinte maneira:

$$\begin{aligned}
\mathbb{P}(Y_t = y_t) &= \sum_{y_{t-1}=0}^{\infty} \mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}) \times \mathbb{P}(Y_{t-1} = y_{t-1}) \\
&= \sum_{y_{t-1}=0}^{\infty} \sum_{k=0}^{\min\{y_t, y_{t-1}\}} \binom{y_{t-1}}{k} \alpha^k (1-\alpha)^{y_{t-1}-k} \\
&\quad \times \left[\pi \mathbb{I}_{\{y_t-k=0\}} + (1-\pi) \mathbb{P}(V_t = y_t - k) \right] \mathbb{P}(Y_{t-1} = y_{t-1}). \quad (3.13)
\end{aligned}$$

Conforme mencionado anteriormente, a distribuição marginal do processo ZI-INAR(1) pode ser representada pela soma infinita de variáveis aleatórias seguindo distribuição ZI. Na seção a seguir apresentamos alguns resultados referentes a distribuição de zeros nos processos ZI-INAR(1).

3.3 DISTRIBUIÇÃO DE ZEROS

O estudo do processo ZI-INAR(1) é motivado pela presença de zeros em séries temporais de valores inteiros. Jazi, Jones e Lai (2012) estudaram o comportamento de zeros em processos INAR(1) com inovações seguindo distribuição ZIP por meio da probabilidade marginal de zero e do valor esperado do *número de zeros consecutivos até o primeiro valor diferente de zero*. Os autores mostraram que a probabilidade marginal de zero no processo com inovações seguindo distribuições ZIP é maior do que no processo com inovações seguindo distribuições Poisson. Sendo assim, na prática, uma maneira simples de verificar se o processo INAR(1) com inovações seguindo distribuição Poisson é adequado para modelar a série, é estimando a probabilidade marginal de zeros do processo e comparando tal probabilidade estimada com a proporção de zeros observada na série temporal. Evidentemente, se a probabilidade estimada de zeros no processo INAR(1) com inovações seguindo distribuição Poisson for menor que a proporção de zeros observada na série temporal, há indícios de que o modelo não é adequado e que um processo com inovações ZIP deve ser considerado.

A seguir, estendemos o estudo do comportamento de zeros para o processo ZI-INAR(1).

Probabilidade de zero

A probabilidade de zero no processo ZI-INAR(1) é dada por:

$$\mathbb{P}(Y_t = 0) = G_{Y_t}(0) = \prod_{j=0}^{\infty} \left[\pi + (1 - \pi) G_{U_t}(1 - \alpha^j | \boldsymbol{\lambda}) \right], \quad (3.14)$$

em que $G_{U_t}(\cdot | \boldsymbol{\lambda})$ é a fgp da v.a. U_t com fp $h_{U_t}(\cdot | \boldsymbol{\lambda})$. Em particular,

- Se $V_t \sim \text{ZIP}(\pi, \lambda)$, então:

$$\mathbb{P}(Y_t = 0) = G_{Y_t}(0) = \prod_{j=0}^{\infty} \left[\pi + (1 - \pi) e^{-\alpha^j \lambda} \right].$$

- Se $V_t \sim \text{ZINB}(\pi, \mu, \phi)$, então:

$$\mathbb{P}(Y_t = 0) = G_{Y_t}(0) = \prod_{j=0}^{\infty} \left[\pi + (1 - \pi) \left(\frac{\phi}{\phi + \alpha^j \mu} \right)^{\phi} \right].$$

- Se $V_t \sim \text{ZIPIG}(\pi, \mu, \phi)$, então:

$$\mathbb{P}(Y_t = 0) = G_{Y_t}(0) = \prod_{j=0}^{\infty} \left[\pi + (1 - \pi) \exp \left\{ \phi \left[1 - \left(1 + \frac{2\alpha^j \mu}{\phi} \right)^{1/2} \right] \right\} \right].$$

Uma maneira de analisar em situações práticas se um processo com inovações ZI deve ser considerado, é utilizando o fato que a probabilidade de zero no processo ZI-INAR(1) é maior ou igual do que no processo INAR(1) com inovações $\{U_t\}_{t \in \mathbb{Z}}$, i.e, com inovações não seguindo distribuições ZI. De fato, reescrevendo a Equação (3.14) temos que a probabilidade de zero no processo ZI-INAR(1) é:

$$\mathbb{P}(Y_t = 0) = \prod_{j=0}^{\infty} \left[G_{U_t}(1 - \alpha^j | \boldsymbol{\lambda}) + \pi(1 - G_{U_t}(1 - \alpha^j | \boldsymbol{\lambda})) \right] \geq \prod_{j=0}^{\infty} G_{U_t}(1 - \alpha^j | \boldsymbol{\lambda}),$$

que é a probabilidade de zero no processo INAR(1) com inovações não pertencente à classe das distribuições ZI, definida por $\mathbb{P}_0(Y_t = 0) = \prod_{j=0}^{\infty} G_{U_t}(1 - \alpha^j | \boldsymbol{\lambda})$. Assim, se ajustarmos um processo INAR(1) com inovações não pertencente à classe das distribuições ZI, e a estimativa de $\mathbb{P}_0(Y_t = 0)$ for menor que a proporção de zeros observada na série temporal, temos indícios que o modelo não é adequado, e que uma distribuição ZI deve ser considerada para a inovação.

Valor esperado do número de zeros consecutivos

A probabilidade de transição $\mathbb{P}(Y_t = 0 | Y_{t-1} = k)$, com $k \in \mathbb{N}^+$, é dada por:

$$\begin{aligned} \mathbb{P}(Y_t = 0 | Y_{t-1} = k) &= \mathbb{P}(\alpha \circ Y_{t-1} = 0 | Y_{t-1} = k) \times \mathbb{P}(V_t = 0) \\ &= \begin{cases} \pi + (1 - \pi)h_{U_t}(0|\boldsymbol{\lambda}), & \text{se } k = 0 \\ (1 - \alpha)^k (\pi + (1 - \pi)h_{U_t}(0|\boldsymbol{\lambda})), & \text{se } k = 1, 2, \dots \end{cases} \end{aligned}$$

No entanto, a probabilidade de transição de “0” para um valor diferente de “0”, denotado por δ , é dada por:

$$\delta = \mathbb{P}(Y_t \neq 0 | Y_{t-1} = 0) = (1 - \pi)(1 - h_{U_t}(0|\boldsymbol{\lambda})). \quad (3.15)$$

Como proposto por Jazi, Jones e Lai (2012), definimos a v.a. N como o “*número de zeros consecutivos até o primeiro valor diferente de zero*”, que representa o número de zeros entre dois valores diferentes de zero. Assim, a v.a. N segue uma distribuição Geométrica com fp:

$$\mathbb{P}(N = k) = \delta(1 - \delta)^{k-1},$$

em que $k \in \mathbb{N}^+$ e δ é definido em (3.15). Assim, o valor esperado da v.a. N é dado por:

$$E(N) = \frac{1}{(1 - \pi)(1 - h_{U_t}(0|\boldsymbol{\lambda}))}. \quad (3.16)$$

Na prática, podemos analisar se o processo INAR(1) com inovações $\{U_t\}_{t \in \mathbb{Z}}$, i.e., com inovações que não seguem distribuições ZI, é adequado por meio do comportamento dos zeros observados na série temporal. Uma maneira de realizar essa análise é utilizando a média do “*número de zeros consecutivos até o primeiro valor diferente de zero*” observado na série, denotado por $\bar{N}$, e dado por:

$$\bar{N} = \frac{\sum_{i=1}^m N_i}{m},$$

em que N_i é o número de elementos da i -ésima sequência de zeros consecutivos da série temporal, e m é o número total de sequências de zeros na série. Note que o valor esperado da v.a. N dado em (3.16) é maior no processo ZI-INAR(1) quando comparado ao processo INAR(1) com inovações $\{U_t\}_{t \in \mathbb{Z}}$. De fato, no processo ZI-INAR(1), temos:

$$E(N) = \frac{1}{(1 - \pi)(1 - h_{U_t}(0|\boldsymbol{\lambda}))} > \frac{1}{1 - h_{U_t}(0|\boldsymbol{\lambda})},$$

que corresponde ao valor esperado do “*número de zeros consecutivos até o primeiro valor diferente de zero*” em um processo INAR(1) com inovações não pertencente à classe das distribuições ZI, definido por $E_0(N) = 1/(1 - h_{U_t}(0|\boldsymbol{\lambda}))$. Assim, se ajustarmos um processo INAR(1) com inovações não pertencente a classe das distribuições ZI, e a estimativa de $E_0(N)$ for menor que $\bar{N}$, temos indícios de que o modelo não é adequado, e que uma distribuição ZI deve ser considerada para a inovação.

3.4 FUNÇÃO DE VEROSSIMILHANÇA

Seja $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ uma realização do processo ZI-INAR(1). A função de verossimilhança de $\boldsymbol{\theta} = (\alpha, \pi, \boldsymbol{\lambda})^\top$, dado $\mathbf{y}$ é definida por:

$$\begin{aligned}
 L(\boldsymbol{\theta}|\mathbf{y}) &= \mathbb{P}(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) \\
 &= \mathbb{P}(Y_1 = y_1) \mathbb{P}(Y_2 = y_2 | Y_1 = y_1) \dots \mathbb{P}(Y_n = y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) \\
 &= \mathbb{P}(Y_1 = y_1) \prod_{t=2}^n \mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}) \\
 &= \mathbb{P}(Y_1 = y_1) \prod_{t=2}^n \left\{ \sum_{k=0}^{\min\{y_{t-1}, y_t\}} \binom{y_{t-1}}{k} \alpha^k (1 - \alpha)^{y_{t-1}-k} \right. \\
 &\quad \left. \times \left(\pi \mathbb{I}_{\{y_t=k\}} + (1 - \pi) h_{U_t}(y_t - k | \boldsymbol{\lambda}) \right) \right\}. \tag{3.17}
 \end{aligned}$$

Note que a função (3.17) é expressa em termos da probabilidade $\mathbb{P}(Y_1 = y_1)$, cuja expressão depende de uma soma infinita (veja Equação (3.13)). Para contornar este problema, utilizamos a função de verossimilhança condicional, que é obtida supondo que $\mathbb{P}(Y_1 = y_1)$ é uma constante fixa, que não depende dos parâmetros. Um estudo de simulação realizado por Jazi, Jones e Lai (2012) mostrou que utilizar esta função em processos com inovações ZIP não implica em uma perda significativa em termos de viés e erro quadrático médio nas estimativas dos parâmetros. Para o processo ZI-INAR(1), a função de log-verossimilhança condicional é dada por:

$$\begin{aligned}
 \ell(\boldsymbol{\theta}|\mathbf{y}) &= \log \{\mathbb{P}(Y_t = y_t)\} + \sum_{t=2}^n \log \{\mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1})\} \\
 &\propto \sum_{t=2}^n \log \left\{ \sum_{k=0}^{\min\{y_{t-1}, y_t\}} \binom{y_{t-1}}{k} \alpha^k (1 - \alpha)^{y_{t-1}-k} \left(\pi \mathbb{I}_{\{y_t=k\}} + (1 - \pi) h_{U_t}(y_t - k | \boldsymbol{\lambda}) \right) \right\}. \tag{3.18}
 \end{aligned}$$

A maximização da função (3.18) não é simples de ser obtida. Sendo assim, para estimar os parâmetros do processo, propomos a metodologia do algoritmo EM, que

apresenta propriedades das quais garantem a convergência monótona para um valor estacionário da função de log-verossimilhança (MCLACHLAN; KRISHNAN, 2008, Capítulo 3). Até o presente momento, não encontramos na literatura estudos que propõem o algoritmo EM para processos INAR.

3.5 ESTIMAÇÃO DOS PARÂMETROS VIA ALGORITMO EM

Nesta seção, desenvolvemos o algoritmo EM para estimação dos parâmetros do processo ZI-INAR(1). Conforme mencionado na Seção 2.4, para desenvolver o algoritmo, será necessário definir as variáveis latentes e em seguida, obter de forma iterativa, no passo E, o valor esperado da função de log-verossimilhança completa (a qual considera a presença das variáveis latentes), e no passo M, a maximização desta função.

Assim, como sugerido por Hall (2000) e Neal e Rao (2007), definimos as variáveis latentes $\mathbf{W} = (W_2, \dots, W_n)^\top$, $\mathbf{S} = (S_2, \dots, S_n)^\top$ e $\mathbf{V} = (V_2, \dots, V_n)^\top$, em que, para todo $t = 2, \dots, n$:

- S_t é uma variável latente, definida por $S_t = \alpha \circ Y_{t-1}$, em que $S_t|Y_{t-1} = y_{t-1} \sim \text{Bin}(y_{t-1}; \alpha)$, se $y_{t-1} > 0$, e $S_t|Y_{t-1} = y_{t-1}$ segue uma distribuição degenerada em zero, se $y_{t-1} = 0$.
- W_t é uma variável latente dicotômica, que assume o valor “1”, com probabilidade π , se a inovação V_t é obtida a partir do estado zero, e “0”, se V_t é obtida a partir da distribuição $h_{U_t}(\cdot|\boldsymbol{\lambda})$, isto é:

$$W_t = \begin{cases} 1, & \text{se } V_t \text{ surgiu do estado zero;} \\ 0, & \text{se } V_t \text{ surgiu da distribuição } h_{U_t}(v_t|\boldsymbol{\lambda}), \end{cases}$$

em que $\mathbb{P}(W_t = 1) = \pi$.

- $V_t = Y_t - S_t$, em que $V_t|W_t = w_t \sim h_{U_t}(v_t|\boldsymbol{\lambda})$, se $W_t = 0$, e $V_t|W_t = w_t$ segue uma distribuição degenerada em zero, se $W_t = 1$.

Definindo $\mathbf{Y}_c = (\mathbf{Y}, \mathbf{W}, \mathbf{S}, \mathbf{V})$ como o vetor dos dados completos, em que $\mathbf{Y}$ e $\{\mathbf{W}, \mathbf{S}, \mathbf{V}\}$ representam os dados observados e os dados faltantes, respectivamente, com $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$, então a função de probabilidade conjunta condicional $\mathbb{P}(\mathbf{Y}_{c_t} = \mathbf{y}_{c_t} | Y_{t-1} =$

y_{t-1}) é dada por:

$$\begin{aligned}
\mathbb{P}(\mathbf{Y}_{c_t} = \mathbf{y}_{c_t} | Y_{t-1} = y_{t-1}) &= \mathbb{P}(Y_t = y_t, W_t = w_t, S_t = s_t, V_t = v_t | Y_{t-1} = y_{t-1}) \\
&= \mathbb{P}(W_t = w_t, S_t = s_t, V_t = v_t | Y_{t-1} = y_{t-1}) \\
&\times \mathbb{P}(Y_t = y_t | W_t = w_t, S_t = s_t, V_t = v_t, Y_{t-1} = y_{t-1}).
\end{aligned} \tag{3.19}$$

Pela definição do processo, temos que $y_t = s_t + v_t$, o que implica que $\mathbb{P}(Y_t = y_t | W_t = w_t, S_t = s_t, V_t = v_t, Y_{t-1} = y_{t-1}) = 1$. Portanto,

$$\begin{aligned}
\mathbb{P}(\mathbf{Y}_{c_t} = \mathbf{y}_{c_t} | Y_{t-1} = y_{t-1}) &= \mathbb{P}(W_t = w_t, S_t = s_t, V_t = v_t | Y_{t-1} = y_{t-1}) \\
&= \mathbb{P}(W_t = w_t, S_t = s_t, V_t = y_t - s_t | Y_{t-1} = y_{t-1}) \\
&= \mathbb{P}(S_t = s_t | Y_{t-1} = y_{t-1}) \times \mathbb{P}(W_t = w_t, V_t = y_t - s_t | S_t = s_t, Y_{t-1} = y_{t-1}) \\
&= \mathbb{P}(S_t = s_t | Y_{t-1} = y_{t-1}) \times \mathbb{P}(W_t = w_t, V_t = y_t - s_t) \\
&= \binom{y_{t-1}}{s_t} \alpha^{s_t} (1 - \alpha)^{y_{t-1} - s_t} \pi^{w_t} [(1 - \pi) h_{U_t}(y_t - s_t)]^{1 - w_t},
\end{aligned} \tag{3.20}$$

em que (3.20) é obtido utilizando o fato que (V_t, W_t) é independente de (S_t, Y_{t-1}) . Assim, a função de verossimilhança condicional completa é dada por:

$$\begin{aligned}
\mathbb{L}_c(\boldsymbol{\theta} | \mathbf{y}_c) &= \mathbb{P}(Y_1 = y_1) \prod_{t=2}^n \mathbb{P}(Y_t = y_t, S_t = s_t, W_t = w_t, V_t = v_t | Y_{t-1} = y_{t-1}) \\
&\propto \prod_{t=2}^n \left\{ \binom{y_{t-1}}{s_t} \alpha^{s_t} (1 - \alpha)^{y_{t-1} - s_t} \pi^{w_t} (1 - \pi)^{1 - w_t} [h_{U_t}(y_t - s_t | \boldsymbol{\lambda})]^{1 - w_t} \right\}.
\end{aligned}$$

A função de log-verossimilhança condicional completa, denotada por $\ell_c(\boldsymbol{\theta} | \mathbf{y}_c)$, é dada por:

$$\begin{aligned}
\ell_c(\boldsymbol{\theta} | \mathbf{y}_c) &= \log(\mathbb{L}_c(\boldsymbol{\theta} | \mathbf{y}_c)) \\
&\propto \sum_{t=2}^n s_t \log(\alpha) + \sum_{t=2}^n (y_{t-1} - s_t) \log(1 - \alpha) \\
&+ \sum_{t=2}^n w_t \log(\pi) + \sum_{t=2}^n (1 - w_t) \log(1 - \pi) \\
&+ \sum_{t=2}^n (1 - w_t) \log h_{U_t}(y_t - s_t | \boldsymbol{\lambda}).
\end{aligned} \tag{3.21}$$

O algoritmo EM é desenvolvido em dois passos:

- **Passo E:** Seja $\hat{\boldsymbol{\theta}}^{(k)}$ a estimativa de $\boldsymbol{\theta}$ para a k -ésima iteração. Determinar a função $Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)})$ definida por:

$$Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}) = E\left(\ell_c(\boldsymbol{\theta} | \mathbf{y}_c) | \mathbf{Y} = \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}\right).$$

- **Passo M:** Maximizar $Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ com relação a $\boldsymbol{\theta}$, obtendo $\hat{\boldsymbol{\theta}}^{(k+1)}$.

Observe que a função $Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ é completamente determinada pelo conhecimento das seguintes esperanças:

$$\hat{s}_t^{(k)} = E\left(S_t|Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right), \quad \hat{w}_t^{(k)} = E\left(W_t|Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right),$$

$$Q_t^*\left(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)}\right) = E\left((1 - W_t) \log h_{U_t}(y_t - S_t) | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)$$

Assim, a função $Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ pode ser expressa como:

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) &\propto \sum_{t=2}^n \hat{w}_t^{(k)} \log(\pi) + \sum_{t=2}^n (1 - \hat{w}_t^{(k)}) \log(1 - \pi) \\ &+ \sum_{t=2}^n \hat{s}_t^{(k)} \log(\alpha) + \sum_{t=2}^n (y_{t-1} - \hat{s}_t^{(k)}) \log(1 - \alpha) + \sum_{t=2}^n Q_t^*(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)}), \end{aligned}$$

Em cada passo, a esperança condicional $\hat{s}_t^{(k)}$ pode ser obtida utilizando o seguinte resultado:

$$\begin{aligned} \mathbb{P}\left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \frac{\mathbb{P}\left(S_t = s_t, Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\ &= \frac{\mathbb{P}\left(S_t = s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(Y_t = y_t | S_t = s_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\ &= \frac{\binom{y_{t-1}}{s_t} (\hat{\alpha}^{(k)})^{s_t} (1 - \hat{\alpha}^{(k)})^{y_{t-1} - s_t} \times \mathbb{P}\left(V_t = y_t - s_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \end{aligned} \quad (3.22)$$

$$\begin{aligned} &= \frac{\binom{y_{t-1}}{s_t} (\hat{\alpha}^{(k)})^{s_t} (1 - \hat{\alpha}^{(k)})^{y_{t-1} - s_t} \left(\hat{\pi}^{(k)} \mathbb{I}_{\{y_t = s_t\}} + (1 - \hat{\pi}^{(k)}) h_{U_t}(y_t - s_t | \hat{\boldsymbol{\lambda}}^{(k)}) \right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}, \end{aligned} \quad (3.23)$$

em que a probabilidade no numerador da expressão (3.22) é consequência de:

$$\begin{aligned} \mathbb{P}(Y_t = y_t | S_t = s_t, Y_{t-1} = y_{t-1}) &= \mathbb{P}(S_t + V_t = y_t | S_t = s_t, Y_{t-1} = y_{t-1}) \\ &= \mathbb{P}(V_t = y_t - s_t | Y_{t-1} = y_{t-1}), \end{aligned} \quad (3.24)$$

em que (3.24) é obtido pela independência das v.a.'s S_t e V_t . Assim, a quantidade $\hat{s}_t^{(k)}$ é obtida por:

$$\begin{aligned}
\hat{s}_t^{(k)} &= E\left(S_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} s_t \mathbb{P}\left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \frac{\sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} s_t \left[\binom{y_{t-1}}{s_t} \left(\hat{\alpha}^{(k)}\right)^{s_t} \left(1 - \hat{\alpha}^{(k)}\right)^{y_{t-1}-s_t} \left(\hat{\pi}^{(k)} \mathbb{I}_{\{y_t=s_t\}} + (1 - \hat{\pi}^{(k)}) h_{U_t}\left(y_t - s_t | \hat{\boldsymbol{\lambda}}^{(k)}\right) \right) \right]}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)},
\end{aligned} \tag{3.25}$$

em que $\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)$ é definida na Equação (3.12).

Por outro lado, em cada passo, $\hat{w}_t^{(k)}$ é obtido como:

$$\begin{aligned}
\hat{w}_t^{(k)} &= E\left(W_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \mathbb{P}\left(W_t = 1 | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \frac{\mathbb{P}\left(W_t = 1, Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(W_t = 1 | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(Y_t = y_t | W_t = 1, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(W_t = 1 | \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(S_t + V_t = y_t | W_t = 1, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}
\end{aligned} \tag{3.26}$$

$$= \frac{\hat{\pi}^{(k)} \times \mathbb{P}\left(S_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \tag{3.27}$$

$$= \frac{\hat{\pi}^{(k)} \binom{y_{t-1}}{y_t} \left(\hat{\alpha}^{(k)}\right)^{s_t} \left(1 - \hat{\alpha}^{(k)}\right)^{y_{t-1}-y_t} \mathbb{I}_{\{y_t \leq y_{t-1}\}}}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}, \tag{3.28}$$

em que (3.26) é obtido utilizando a independência das v.a.'s W_t e Y_{t-1} , e (3.27) é obtido pelo fato de S_t e (V_t, W_t) serem independentes e, dado $W_t = 1$, a v.a. V_t é degenerada em zero.

Sendo assim, nosso algoritmo EM para o processo ZI-INAR(1) pode ser resumido como:

- **Passo E:** Dado $\hat{\boldsymbol{\theta}}^{(k)}$, para $t = 2, 3, \dots$, calculamos $\hat{s}_t^{(k)}$ e $\hat{w}_t^{(k)}$, dado nas equações (3.25) e (3.28), respectivamente, e $Q_t^*(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)})$, o qual mostraremos sua expressão para os casos em que as inovações seguem distribuição ZIP, ZINB e ZIPIG.
- **Passo M:** Atualizamos $\hat{\boldsymbol{\theta}}^{(k)}$ pela maximização $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ sobre $\boldsymbol{\theta}$, o que implica nas seguintes expressões:

$$\hat{\alpha}^{(k+1)} = \frac{\sum_{t=2}^n \hat{s}_t}{\sum_{t=2}^n y_{t-1}}$$

$$\hat{\pi}^{(k+1)} = \frac{\sum_{t=2}^n \hat{w}_t}{n-1}$$

$$\hat{\boldsymbol{\lambda}}^{(k+1)} = \arg \max_{\boldsymbol{\lambda}} \left\{ \sum_{t=2}^n Q_t^*(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)}) \right\}.$$

Os passos E e M do algoritmo são iterados até que algum critério de convergência seja satisfeito. Neste trabalho, utilizamos o critério de parada baseado na aceleração de Aitken (MCLACHLAN; KRISHNAN, 2008) como critério de convergência. Sendo $\ell^{(k+1)} = \ell(\hat{\boldsymbol{\theta}}^{(k+1)}|\mathbf{y})$, em que $\ell(\boldsymbol{\theta}|\mathbf{y})$ é dado em (3.18), o critério baseado na aceleração de Aitken utiliza o fato que o limite da sequência $\ell^{(1)}, \ell^{(2)}, \dots$, denotado por ℓ_∞ , pode ser aproximado por $\ell_\infty^{(k+1)} = \ell^{(k)} + (\ell^{(k+1)} - \ell^{(k)})/(1 - c^{(k)})$, em que $c^{(k)} = (\ell^{(k+1)} - \ell^{(k)})/(\ell^{(k)} - \ell^{(k-1)})$. Assim como Zeller *et al.* (2019), decidimos utilizar o seguinte critério de convergência: $\|\ell_\infty^{(k+1)} - \ell^{(k+1)}\| < \varepsilon$. Adotamos $\varepsilon = 10^{-5}$.

A seguir, mostramos como obter $\hat{\boldsymbol{\lambda}}^{(k+1)}$ para os casos particulares em que as inovações seguem uma distribuição ZIP, ZINB e ZIPIG. O conhecimento do valor esperado $E\left(B_t S_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)$, denotado por $\widehat{b_t s_t}$, sendo $B_t = 1 - W_t$, será útil para prosseguir com o passo E do algoritmo EM nos três casos particulares mencionados. Assim, este valor esperado é dado por

$$\begin{aligned} \widehat{b_t s_t} &= E\left(B_t S_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\ &= E\left\{E\left(B_t S_t | B_t, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)\right\} \\ &= E\left\{B_t E\left(S_t | B_t, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right\} \\ &= \mathbb{P}\left(B_t = 1 | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) E\left(S_t | B_t = 1, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right), \end{aligned} \tag{3.29}$$

em que

$$\begin{aligned}
\mathbb{P}\left(B_t = 1 | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \frac{\mathbb{P}\left(W_t = 0, Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(W_t = 0 | \hat{\boldsymbol{\theta}}^{(k)}\right) \mathbb{P}\left(Y_t = y_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{(1 - \hat{\pi}^{(k)}) \mathbb{P}\left(Y_t = y_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}. \tag{3.30}
\end{aligned}$$

Para calcular o valor esperado na expressão (3.29), temos que:

$$\begin{aligned}
\mathbb{P}\left(S_t = s_t | W_t = 0, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \\
&= \frac{\mathbb{P}\left(S_t = s_t, Y_t = y_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(S_t = s_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(Y_t = y_t | S_t = s_t, W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(S_t = s_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(V_t = y_t - s_t | W_t = 0, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}.
\end{aligned}$$

Logo:

$$\begin{aligned}
E\left(S_t | B_t = 1, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \\
&= \frac{\sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} s_t \left[\mathbb{P}\left(S_t = s_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(V_t = y_t - s_t | W_t = 0, \hat{\boldsymbol{\theta}}^{(k)}\right) \right]}{\mathbb{P}\left(Y_t = y_t | W_t = 0, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)}\right)}. \tag{3.31}
\end{aligned}$$

Note que, podemos simplificar a expressão (3.29) utilizando os resultados (3.30) e (3.31), o que implica que:

$$\begin{aligned}
\widehat{b_t s_t} &= \mathbb{P}\left(B_t = 1 | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)}\right) \times E\left(S_t | B_t = 1, Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \frac{\left(1 - \widehat{\pi}^{(k)}\right) \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} s_t \mathbb{P}\left(S_t = s_t | Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(V_t = y_t - s_t | W_t = 0, \widehat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\left(1 - \widehat{\pi}^{(k)}\right) \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} s_t \binom{y_{t-1}}{s_t} \left(\widehat{\alpha}^{(k)}\right)^{s_t} \left(1 - \widehat{\alpha}^{(k)}\right)^{y_{t-1} - s_t} h_{U_t}\left(y_t - s_t | \widehat{\boldsymbol{\lambda}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)}\right)}. \quad (3.32)
\end{aligned}$$

1. Caso ZIP:

Se $V_t \sim \text{ZIP}(\pi, \lambda)$, então $\boldsymbol{\lambda} = \lambda$ e $h_{U_t}(\cdot | \boldsymbol{\lambda})$ é a fp de uma v.a. Poisson(λ). Assim, temos que:

$$(1 - w_t) \log \{h_{U_t}(y_t - s_t | \boldsymbol{\lambda})\} \propto -(1 - w_t)\lambda + (1 - w_t)(y_t - s_t) \log(\lambda).$$

Logo,

$$\sum_{t=2}^n Q_t^*(\mu | \boldsymbol{\theta}^{(k)}) \propto -\lambda \sum_{t=2}^n (1 - \widehat{w}_t^{(k)}) + \log(\lambda) \sum_{t=2}^n (1 - \widehat{w}_t^{(k)}) y_t + \log(\lambda) \sum_{t=2}^n \widehat{b_t s_t}^{(k)}. \quad (3.33)$$

Maximizando a expressão (3.33) com relação a λ , temos que:

$$\widehat{\lambda}^{(k+1)} = \frac{\sum_{t=2}^n (1 - \widehat{w}_t^{(k)}) y_t - \sum_{t=2}^n \widehat{b_t s_t}^{(k)}}{\sum_{t=2}^n (1 - \widehat{w}_t^{(k)})}.$$

Note que a atualização de $\widehat{\lambda}^{(k)}$ pertence ao espaço paramétrico de λ . De fato, temos que:

$$\begin{aligned}
\sum_{t=2}^n (1 - \widehat{w}_t^{(k)}) y_t - \sum_{t=2}^n \widehat{b_t s_t}^{(k)} &= \sum_{t=2}^n \left(1 - E\left(W_t | \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right)\right) y_t - \sum_{t=2}^n E\left(B_t S_t | \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \sum_{t=2}^n E\left(B_t Y_t | \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right) - \sum_{t=2}^n E\left(B_t S_t | \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \sum_{t=2}^n E\left(B_t (Y_t - S_t) | \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \sum_{t=2}^n E\left(B_t V_t | \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right) > 0.
\end{aligned}$$

2. Caso ZINB:

Se $V_t \sim \text{ZINB}(\pi, \mu, \phi)$, então $\boldsymbol{\lambda} = (\mu, \phi)$ e $h_{U_t}(\cdot | \boldsymbol{\lambda})$ é a fp da v.a. $\text{NB}(\mu, \phi)$. Assim, temos que:

$$(1 - w_t) \log \{h_{U_t}(y_t - s_t | \boldsymbol{\lambda})\} \propto (1 - w_t) \left\{ \log \Gamma(y_t - s_t + \phi) - \log(\Gamma(\phi)) \right. \\ \left. + (y_t - s_t) [\log(\mu) - \log(\mu + \phi)] + \phi [\log(\phi) - \log(\mu + \phi)] \right\}.$$

Logo:

$$\begin{aligned} \sum_{t=2}^n Q_t^*(\mu, \phi | \boldsymbol{\theta}^{(k)}) &\propto \sum_{t=2}^n \widehat{g_t(\phi)}^{(k)} + [\log(\mu) - \log(\mu + \phi)] \\ &\times \left(\sum_{t=2}^n (1 - \widehat{w_t}^{(k)}) \widehat{y_t} - \sum_{t=2}^n \widehat{b_t s_t}^{(k)} \right) \\ &+ [-\log(\Gamma(\phi)) + \phi [\log(\phi) - \log(\mu + \phi)]] \sum_{t=2}^n \left(1 - \widehat{w_t}^{(k)} \right), \quad (3.34) \end{aligned}$$

em que $\widehat{g_t(\phi)}^{(k)} = E \left(B_t \log \Gamma(y_t - S_t + \phi) | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)} \right)$, que é obtido analogamente a (3.32), o que implica em:

$$\widehat{g_t(\phi)}^{(k)} = \frac{\left(1 - \widehat{\pi}^{(k)} \right) \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \log \Gamma(y_t - s_t + \phi) \mathbb{P} \left(S_t = s_t | Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)} \right) h \left(y_t - s_t | \widehat{\boldsymbol{\theta}}^{(k)} \right)}{\mathbb{P} \left(Y_t = y_t | Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)} \right)}.$$

No passo M do algoritmo, maximiza-se a expressão (3.34) com respeito a μ e ϕ . No entanto, tal maximização não pode ser obtida analiticamente. Uma maneira de contornar este problema, é utilizando o algoritmo ECM proposto por Meng e Rubin (1993), o qual substitui o passo M por uma sequência de passos de maximização condicional (CM). Sendo assim, o passo CM do algoritmo para o processo ZINB-INAR(1) segue os seguintes passos:

1. Atualize $\widehat{\mu}^{(k)}$ pela maximização de $\sum_{t=2}^n Q_t^* \left(\mu, \widehat{\phi}^{(k)} | \widehat{\boldsymbol{\theta}}^{(k)} \right)$ com relação a μ .
2. Atualize $\widehat{\phi}^{(k)}$ pela maximização de $\sum_{t=2}^n Q_t^* \left(\widehat{\mu}^{(k+1)}, \phi | \widehat{\boldsymbol{\theta}}^{(k)} \right)$ com relação a ϕ .

Assim, a atualização de $\widehat{\mu}^{(k)}$ e $\widehat{\phi}^{(k)}$ são dadas por:

$$\begin{aligned} \widehat{\mu}^{(k+1)} &= \frac{\sum_{t=2}^n (1 - \widehat{w_t}^{(k)}) \widehat{y_t} - \sum_{t=2}^n \widehat{b_t s_t}^{(k)}}{\sum_{t=2}^n (1 - \widehat{w_t}^{(k)})} \\ \widehat{\phi}^{(k+1)} &= \arg \max_{\phi} \left\{ \sum_{t=2}^n Q_t^* \left(\widehat{\mu}^{(k+1)}, \phi | \widehat{\boldsymbol{\theta}}^{(k)} \right) \right\}. \end{aligned}$$

Não há fórmula explícita para a solução de $\widehat{\phi}^{(k+1)}$. Portanto, utilizamos a rotina “*optim*” do software R (R Core Team, 2019) para maximizar $\sum_{t=2}^n Q_t^* \left(\widehat{\mu}^{(k+1)}, \phi | \widehat{\boldsymbol{\theta}}^{(k)} \right)$ com relação

a ϕ .

3. Caso ZIPIG:

Se $V_t \sim \text{ZIPIG}(\pi, \mu, \phi)$, então $\lambda = (\mu, \phi)$ e a distribuição de V_t tem a seguinte forma estocástica:

- $(V_t|Z_t = z_t, W_t = w_t) \sim \text{Poisson}(\mu z_t)$, se $w_t = 0$, e $(V_t|Z_t = z_t, W_t = w_t)$ segue uma v.a. degenerada em zero, se $w_t = 1$, $t = 2, 3, \dots$
- $Z_t|W_t = 0 \sim \text{IG}(1, \phi)$, $t = 2, 3, \dots$
- $W_t \sim \text{Bin}(1, \pi)$, $t = 2, 3, \dots$

Para o caso ZIPIG, a maximização da função $Q_t^*(\mu, \phi|\hat{\theta}^{(k)})$ não é simples de ser obtida devido a função Bessel de terceira ordem em sua expressão. Sendo assim, para desenvolver o algoritmo EM, iremos propor outra variável latente Z_t , em que o vetor de dados completos, neste caso, será $\mathbf{Y}_c = (\mathbf{Y}, \mathbf{W}, \mathbf{S}, \mathbf{Z})$. Então, utilizando o mesmo argumento que (3.19), a fdp conjunta condicional $f(\mathbf{y}_{c_t}|y_{t-1})$ é dada por:

$$f(\mathbf{y}_{c_t}|y_{t-1}) = f(y_t, s_t, w_t, v_t, z_t|y_{t-1}) = f(y_t, s_t, w_t, z_t|y_{t-1}).$$

Logo:

$$\begin{aligned} f(\mathbf{y}_{c_t}|y_{t-1}) &= f(y_t, s_t, w_t, z_t|y_{t-1}) \\ &= f(y_t, w_t, z_t|s_t, y_{t-1}) \times \mathbb{P}(S_t = s_t|Y_{t-1} = y_{t-1}) \\ &= f(w_t, z_t) \times \mathbb{P}(Y_t = y_t|S_t = s_t, W_t = w_t, Z_t = z_t, Y_{t-1} = y_{t-1}) \end{aligned} \quad (3.35)$$

$$\begin{aligned} &\times \mathbb{P}(S_t = s_t|Y_{t-1} = y_{t-1}) \\ &= f(z_t|w_t) \mathbb{P}(W_t = w_t) \times \mathbb{P}(V_t = y_t - s_t|W_t = w_t, Z_t = z_t) \end{aligned} \quad (3.36)$$

$$\begin{aligned} &\times \mathbb{P}(S_t = s_t|Y_{t-1} = y_{t-1}) \\ &= \left\{ \frac{\phi^{1/2}}{\sqrt{2\pi}} z_t^{-3/2} \exp \left\{ -\frac{\phi(z_t - 1)^2}{2z_t} \right\} \right\}^{1-w_t} \pi^{w_t} (1 - \pi)^{1-w_t} \left[\frac{e^{-\mu z_t} (\mu z_t)^{y_t - s_t}}{(y_t - s_t)!} \right]^{1-w_t} \\ &\times \binom{y_{t-1}}{s_t} \alpha^{s_t} (1 - \alpha)^{y_{t-1} - s_t}, \end{aligned}$$

em que (3.35) é obtido considerando que o vetor (W_t, Z_t) é independente de (S_t, Y_{t-1}) , e (3.36) é obtido analogamente à expressão (3.24).

Assim, a função de verossimilhança condicional completa é dada por:

$$\begin{aligned}\mathbb{L}_c(\boldsymbol{\theta}|\mathbf{y}_c) &\propto \prod_{t=2}^n \left\{ \pi^{w_t} (1-\pi)^{1-w_t} \binom{y_t-1}{s_t} \alpha^{s_t} (1-\alpha)^{y_t-1-s_t} \left[\frac{e^{-\mu z_t} (\mu z_t)^{y_t-s_t}}{(y_t-s_t)!} \right]^{1-w_t} \right. \\ &\quad \times \left. \left\{ \phi^{1/2} z_t^{-3/2} \exp \left\{ -\frac{\phi}{2} \frac{(z_t-1)^2}{z_t} \right\} \right\}^{1-w_t} \right\}.\end{aligned}$$

A função de log-verossimilhança condicional completa é dada por:

$$\begin{aligned}\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \log(\mathbb{L}_c(\boldsymbol{\theta}|\mathbf{y}_c)) \\ &\propto \sum_{t=2}^n w_t \log(\pi) + \sum_{t=2}^n (1-w_t) \log(1-\pi) \\ &\quad + \sum_{t=2}^n s_t \log(\alpha) + \sum_{t=2}^n (y_t-s_t) \log(1-\alpha) \\ &\quad + \sum_{t=2}^n (1-w_t) \left[-\mu z_t + (y_t-s_t) \log(\mu) - \frac{\phi}{2} z_t - \frac{\phi}{2} z_t^{-1} + \phi + \frac{1}{2} \log(\phi) \right].\end{aligned}$$

Assim, a função $Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ pode ser expressa na forma reduzida como:

$$\begin{aligned}Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) &\propto \sum_{t=2}^n \hat{w}_t^{(k)} \log(\pi) + \sum_{t=2}^n \left(1-\hat{w}_t^{(k)}\right) \log(1-\pi) \\ &\quad + \sum_{t=2}^n \hat{s}_t^{(k)} \log(\alpha) + \sum_{t=2}^n \left(y_t-\hat{s}_t^{(k)}\right) \log(1-\alpha) + \sum_{t=2}^n Q_t^{**}(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)}),\end{aligned}$$

em que:

$$\begin{aligned}Q_t^{**}(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)}) &\propto -\mu \sum_{t=2}^n \widehat{b_t z_t}^{(k)} - \log(\mu) \sum_{t=2}^n \widehat{b_t s_t}^{(k)} + \log(\mu) \sum_{t=2}^n (1-\hat{w}_t^{(k)}) y_t \\ &\quad + \frac{\log(\phi)}{2} \sum_{t=2}^n (1-\hat{w}_t^{(k)}) + \phi \sum_{t=2}^n (1-\hat{w}_t^{(k)}) - \frac{\phi}{2} \sum_{t=2}^n \widehat{b_t z_t}^{(k)} - \frac{\phi}{2} \sum_{t=2}^n \widehat{b_t/z_t}^{(k)}.\end{aligned}$$

Assim, no passo E do algoritmo, obtém-se $\hat{s}_t, \hat{w}_t$ e a função $Q_t^{**}\left(\mu, \phi|\hat{\boldsymbol{\theta}}^{(k)}\right)$, a qual é completamente determinada pelo conhecimento de $\widehat{b_t s_t}^{(k)}, \widehat{b_t z_t}^{(k)}, \widehat{b_t/z_t}^{(k)}$, em que $\widehat{b_t z_t}^{(k)} = E\left(B_t Z_t|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}\right)$ e $\widehat{b_t/z_t}^{(k)} = E\left(B_t Z_t^{-1}|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}\right)$. As esperanças $\widehat{b_t z_t}^{(k)}$ e $\widehat{b_t/z_t}^{(k)}$ podem ser expressas por:

$$\begin{aligned}
\widehat{b_t z_t}^{(k)} &= E \left[E \left(B_t Z_t | S_t, Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\lambda}}^{(k)} \right) \right] \\
&= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \frac{(1 - \widehat{\pi}^{(k)}) (y_t - s_t + 1) h_{U_t} \left(y_t - s_t + 1 | \widehat{\boldsymbol{\lambda}}^{(k)} \right) \mathbb{P} \left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)} \right)}{\widehat{\mu}^{(k)} \left[\widehat{\pi}^{(k)} I_{\{y_t = s_t\}} + (1 - \widehat{\pi}^{(k)}) h_{U_t} \left(y_t - s_t | \widehat{\boldsymbol{\lambda}}^{(k)} \right) \right]}
\end{aligned} \tag{3.37}$$

e

$$\begin{aligned}
\widehat{b_t/z_t}^{(k)} &= E \left[E \left(B_t Z_t^{-1} | S_t, Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\lambda}}^{(k)} \right) \right] \\
&= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} E \left(B_t Z_t^{-1} | V_t = y_t - s_t, \widehat{\boldsymbol{\lambda}}^{(k)} \right) \times \mathbb{P} \left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\lambda}}^{(k)} \right) \\
&= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \left\{ \frac{\widehat{\mu}^{(k)} h \left(y_t - s_t - 1 | \widehat{\boldsymbol{\lambda}}^{(k)} \right) \times \mathbb{P} \left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)} \right)}{(y_t - s_t) h \left(y_t - s_t | \widehat{\boldsymbol{\lambda}}^{(k)} \right)} \mathbb{I}_{\{y_t > s_t\}} \right. \\
&\quad \left. + \frac{\left(\sqrt{\widehat{\phi}^{(k)} (\widehat{\phi}^{(k)} + 2\widehat{\mu}^{(k)})} + 1 \right) (1 - \widehat{\pi}^{(k)}) h \left(0 | \widehat{\boldsymbol{\lambda}}^{(k)} \right) \times \mathbb{P} \left(S_t = y_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)} \right)}{\widehat{\phi}^{(k)} \left[\widehat{\pi}^{(k)} + (1 - \widehat{\pi}^{(k)}) h \left(0 | \widehat{\boldsymbol{\lambda}}^{(k)} \right) \right]} \mathbb{I}_{\{y_t = s_t\}} \right\},
\end{aligned} \tag{3.38}$$

em que $\mathbb{P} \left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \widehat{\boldsymbol{\theta}}^{(k)} \right)$ é dado em (3.23) e a demonstração de (3.37) e (3.38) estão no Apêndice B.

No passo M do algoritmo, substituímos a maximização de $\sum_{t=2}^n Q_t^* \left(\mu, \phi | \widehat{\boldsymbol{\theta}}^{(k)} \right)$ pela maximização de $\sum_{t=2}^n Q_t^{**} \left(\mu, \phi | \widehat{\boldsymbol{\theta}}^{(k)} \right)$. Assim, temos que a atualização de $\widehat{\mu}^{(k)}$ e $\widehat{\phi}^{(k)}$ é obtida por:

$$\begin{aligned}
\widehat{\mu}^{(k+1)} &= \frac{\sum_{t=2}^n (1 - \widehat{w}_t^{(k)}) y_t - \sum_{t=2}^n \widehat{b_t s_t}^{(k)}}{\sum_{t=2}^n \widehat{b_t z_t}^{(k)}}. \\
\widehat{\phi}^{(k+1)} &= \frac{\sum_{t=2}^n (1 - \widehat{w}_t^{(k)})}{\sum_{t=2}^n \widehat{b_t z_t}^{(k)} + \sum_{t=2}^n \widehat{b_t/z_t}^{(k)} - 2 \sum_{t=2}^n (1 - \widehat{w}_t^{(k)})}.
\end{aligned}$$

Em seguida, mostramos os valores iniciais que adotamos para as estimativas de α , π e $\boldsymbol{\lambda}$ na primeira iteração do algoritmo.

Valores iniciais

Para os valores iniciais $\hat{\boldsymbol{\theta}}^{(0)} = (\hat{\alpha}^{(0)}, \hat{\pi}^{(0)}, \hat{\lambda}^{(0)})$, utilizamos as estimativas obtidas pelo método de momentos, igualando a esperança, variância e função de autocovariância do processo à esperança, variância e função de autocovariância amostral, respectivamente (veja Eq. (2.3) e (2.4) – Cap. 2). Assim, temos que:

$$\hat{\alpha}^{(0)} = \hat{\rho}(1) = \frac{\sum_{t=1}^{n-1} (y_t - \bar{y})(y_{t+1} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2},$$

em que $\bar{y} = \sum_{i=1}^n y_i / n$.

Relacionado as estimativas iniciais para π e λ , temos:

- Se $V_t \sim \text{ZIP}(\pi, \lambda)$, então:

$$\hat{\pi}^{(0)} = \frac{(\hat{\delta} - 1)(1 + \hat{\alpha}^{(0)})}{(\hat{\delta} - 1)(1 + \hat{\alpha}^{(0)}) + \bar{y}(1 - \hat{\alpha}^{(0)})} \quad \text{e} \quad \hat{\lambda} = \frac{\bar{y}(1 - \hat{\alpha}^{(0)})}{1 - \hat{\pi}^{(0)}}.$$

- Se $V_t \sim \text{ZINB}(\pi, \lambda)$ ou $V_t \sim \text{ZIPIG}(\pi, \lambda)$, então:

$$\hat{\mu}^{(0)} = \frac{\bar{y}(1 - \hat{\alpha}^{(0)})}{1 - \hat{\pi}^{(0)}} \quad \text{e} \quad \hat{\phi}^{(0)} = \frac{\hat{\mu}^{(0)}}{(\hat{\delta} - 1)(1 + \hat{\alpha}^{(0)}) - \hat{\mu}^{(0)}\hat{\pi}^{(0)}},$$

em que $\hat{\delta} = S^2 / \bar{y}$, com $S^2 = \sum_{t=1}^n (y_t - \bar{y})^2 / (n - 1)$. Uma vez que não há fórmula explícita para a estimativa pelo método de momentos para o parâmetro π , utilizamos $\hat{\pi}^{(0)} = 0,5$, que corresponde ao ponto médio do intervalo (0,1) (espaço paramétrico de π).

A seguir, mostramos como construir intervalos de confiança para os parâmetros por meio de métodos bootstrap.

3.6 INTERVALOS DE CONFIANÇA PARA OS PARÂMETROS

No processo ZI-INAR(1) não é conhecida a distribuição assintótica dos estimadores de máxima verossimilhança. Assim, com o intuito de construir intervalos de confiança para os parâmetros, propomos o procedimento de reamostragem por bootstrap, que utiliza os dados observados para gerar amostras que possibilitam estimar a distribuição dos estimadores.

Os métodos bootstrap, inicialmente propostos por Efron (1979), não são adequados para conjunto de dados que apresentem estrutura de dependência, como em séries

temporais. No contexto de processos INAR, alguns estudos propondo métodos bootstrap têm sido desenvolvidos para construir intervalos de confiança e realizar previsões, dentre os quais podemos citar: Cardinal, Roy e Lambert (1999), Jung e Tremayne (2006), Kim e Park (2008), Kim e Park (2010), Jentsch e Weiß (2019) e Bisaglia e Gerolimetto (2019). Sendo assim, nesta seção apresentamos dois métodos bootstrap para o nosso processo ZI-INAR(1): o bootstrap paramétrico, em que obtemos as amostras gerando valores do modelo ajustado, e um bootstrap não paramétrico, em que as amostras são obtidas por reamostragem dos dados observados.

- *Bootstrap ZI-INAR(1):*

A versão paramétrica do bootstrap para o nosso processo, a qual denotamos por bootstrap ZI-INAR(1), consiste em ajustar um processo ZI-INAR(1) a partir da série observada e posteriormente gerar amostras do processo ajustado. Assim, seja $\mathbf{y} = (y_1, \dots, y_n)$ realizações de um processo ZI-INAR(1), então o bootstrap paramétrico é dado pelas seguintes etapas:

1. Gere as v.a.'s $V_1^*, \dots, V_n^*$, as quais têm distribuição ZI com parâmetros $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\pi}, \hat{\boldsymbol{\lambda}})^\top$, em que $\hat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança obtida pelo algoritmo EM considerando os dados originais.
2. Gere um processo ZI-INAR(1) recursivamente como:

$$\begin{aligned} Y_1^* &= y_1 \\ Y_t^* &= \hat{\alpha} \circ Y_{t-1}^* + V_t^*, \quad t = 2, \dots, n, \end{aligned}$$

em que $\hat{\alpha} \circ Y_{t-1}^* | Y_{t-1}^* = y_{t-1}^*$ segue uma distribuição Binomial com parâmetros y_{t-1}^* e $\hat{\alpha}$, se $y_{t-1}^* > 0$, e $\hat{\alpha} \circ Y_{t-1}^* | Y_{t-1}^* = y_{t-1}^*$ tem distribuição degenerada em zero, se $y_{t-1}^* = 0$.

3. Calcule para a série temporal gerada no passo 2 a estimativa bootstrap $\hat{\boldsymbol{\theta}}^* = (\hat{\alpha}^*, \hat{\pi}^*, \hat{\boldsymbol{\lambda}}^*)^\top$ utilizando o algoritmo EM.
4. Repita os passos (1)-(3), R vezes, obtendo as estimativas bootstrap $\{\hat{\boldsymbol{\theta}}_r^*, r = 1, \dots, R\}$.
5. Utilize a distribuição empírica das estimativas $\hat{\boldsymbol{\theta}}_1^*, \dots, \hat{\boldsymbol{\theta}}_R^*$ como a estimativa da distribuição de $\hat{\boldsymbol{\theta}}$.

- *Bootstrap em blocos circulares:*

Conforme mencionado na Seção 2.5 do Capítulo 2, existem diversas variações do bootstrap em blocos, sendo uma delas o *bootstrap circular em blocos*, em que para

definir os blocos, considera-se que os mesmos estão sobrepostos (isto é, o primeiro bloco se inicia pela observação y_1 , o segundo, por y_2 , e assim por diante) e que a série original não tem um valor inicial e um valor final (isto é, se no bloco há a observação y_n , então o próximo elemento do bloco será y_1 , seguido por y_2 , e assim por diante). Assim, sendo $\mathbf{y} = (y_1, \dots, y_n)$ uma realização de um processo ZI-INAR(1), o bootstrap circular em blocos é dado pelas seguintes etapas:

1. Escolha um tamanho l para os blocos e defina os blocos $B_1, B_2, \dots, B_n$ por:

$$B_i = \begin{cases} (y_i, \dots, y_{i+l-1}), & \text{se } i = 1, \dots, n-l+1; \\ (y_i, \dots, y_n, y_1, \dots, y_{l-(n-i+1)}), & \text{se } i = n-l+2, \dots, n. \end{cases}$$

2. Selecione com reposição $k = n/l$ blocos para gerar a série temporal $\mathbf{y}^* = (y_1^*, \dots, y_{kl}^*)^\top$, em que $y_{1+(i-1)l}^*, y_{2+(i-1)l}^*, \dots, y_{il}^*$ correspondem as observações do i -ésimo bloco selecionado. Caso n/l não seja um número inteiro, selecione $k = \lfloor \frac{n}{l} \rfloor + 1$ blocos, em que $\lfloor a \rfloor$ denota a função maior inteiro, e desconsidere as observações $y_{n+1}^*, \dots, y_{kl}^*$ da série temporal $\mathbf{y}^*$.
3. Calcule para a série temporal $\mathbf{y}^*$ a estimativa bootstrap $\hat{\boldsymbol{\theta}}^* = (\hat{\alpha}^*, \hat{\pi}^*, \hat{\boldsymbol{\lambda}}^*)^\top$ utilizando o algoritmo EM.
4. Repita os passos (1)-(3) R vezes, obtendo as estimativas $\{\hat{\boldsymbol{\theta}}_r^*, r = 1, \dots, R\}$.
5. Utilize a distribuição empírica das estimativas $\hat{\boldsymbol{\theta}}_1^*, \dots, \hat{\boldsymbol{\theta}}_R^*$ como a estimativa da distribuição de $\hat{\boldsymbol{\theta}}$.

Note que os blocos permitem que a estrutura de dependência até a ordem l seja preservada. Para mais detalhes relacionados ao bootstrap em blocos, veja Davison e Hinkley (1997).

A partir das distribuições empíricas obtidas nos passos (1)-(5) para os dois métodos bootstrap, podemos construir intervalos de confiança para os parâmetros. Assim, sendo $\hat{\theta}_i$ a estimativa do parâmetro θ_i do processo ZI-INAR(1), e $\hat{\theta}_{r,i}^*$ a estimativa de θ_i da r -ésima amostra bootstrap, utilizamos os seguintes intervalos:

- *Intervalo de confiança baseada na variância assintótica:*

A variância bootstrap de $\hat{\theta}_i$ é dada por:

$$V_i^* = \frac{1}{R} \sum_{j=1}^R \left(\hat{\theta}_{r,i}^* - \frac{1}{R} \sum_{j=1}^R \hat{\theta}_{r,i}^* \right)^2.$$

Assim, o intervalo assintótico com confiança $1 - \alpha$ para θ_i é dado por:

$$\left[\hat{\theta}_i - u_{1-\alpha/2} \sqrt{V_i^*}, \hat{\theta}_i + u_{1-\alpha/2} \sqrt{V_i^*} \right],$$

em que $u_{1-\alpha/2}$ é o quantil de ordem $1 - \alpha/2$ da distribuição normal padrão.

- *Intervalo de confiança percentílico:*

Definindo $\hat{\theta}_{i(\alpha/2)}^*$ e $\hat{\theta}_{i(1-\alpha/2)}^*$ como os quantis de ordem $\alpha/2$ e $1 - \alpha/2$ da distribuição de $\hat{\theta}_i$ obtida pelo método bootstrap, temos que o intervalo de confiança percentílico com confiança $1 - \alpha$ para θ_i é dado por:

$$\left[\hat{\theta}_{i(\alpha/2)}^*, \hat{\theta}_{i(1-\alpha/2)}^* \right].$$

Além de construir intervalos de confiança para os parâmetros, mostramos na seção a seguir que os métodos bootstrap podem ser utilizados para obter a distribuição de previsão h passos à frente.

3.7 PREVISÃO

Um dos principais objetivos na análise de séries temporais é prever valores futuros. Sendo assim, apresentamos nesta seção métodos de previsões para o processo ZI-INAR(1).

Previsão h passos à frente

Estudos de previsões para séries temporais de valores inteiros com inflação de zeros têm sido pouco explorados na literatura. Jazi, Jones e Lai (2012) sugerem como proposta de previsão a esperança condicional $E(Y_{n+h}|Y_n = y_n)$ em processos INAR(1), com inovações seguindo distribuição ZIP. A previsão baseada na esperança condicional é um dos métodos mais utilizados para realizar previsões em processos autorregressivos contínuos devido ao fato que $E(Y_{n+h}|Y_n = y_n)$ minimiza o erro quadrático médio de previsão, ou seja, minimiza a quantidade $E\left\{\left[\hat{Y}_{n+h} - Y_{n+h}\right]^2 | Y_n = y_n\right\}$ com relação a $\hat{Y}_{n+h}$, que é a previsão h passos à frente.

Sendo assim, propomos a previsão baseada na esperança condicional para a classe dos processos ZI-INAR(1). A estimativa da esperança condicional de Y_{n+h} dado $Y_n = y_n$, denotada por $\hat{y}_{n+h}$, é dada por:

$$\hat{y}_{n+h} = E(Y_{n+h}|Y_n, \hat{\theta}) = \hat{\alpha}^h y_t + (1 - \hat{\pi}) E\left(U_t | \hat{\lambda}\right) \frac{1 - \hat{\alpha}^h}{1 - \hat{\alpha}}, \quad (3.39)$$

em que $E(Y_{n+h}|Y_n, \hat{\theta})$ é obtido pela Equação (3.6).

Para analisar a acurácia da previsão (3.39), utilizamos a medida do Erro Absoluto Médio de Previsão h passos à frente, denotada por $EAMP(h)$, que considera a diferença absoluta entre o valor da previsão e o verdadeiro valor observado na série temporal. Assim, seja $y_1, \dots, y_{n-m}, y_{n-m+1}, \dots, y_n$ uma série temporal de tamanho n dividida em duas partes. Considere que a primeira parte consiste nos $n - m$ valores iniciais da série, utilizados para a estimação dos parâmetros, e a segunda parte, consiste nas m últimas observações, utilizadas para analisar a acurácia das previsões. Sendo $\hat{y}_{t+h}$ a previsão h passos à frente baseada nas observações $y_1, \dots, y_t$. Então, o Erro Absoluto Médio de Previsão h Passos à Frente é dado por:

$$EAMP(h) = \frac{1}{m - h + 1} \sum_{t=n-m}^{n-h} |\hat{y}_{t+h} - y_{t+h}|. \quad (3.40)$$

Distribuição de previsão h passos à frente

No contexto de previsões para processos autorregressivos de valores inteiros, há poucos estudos que utilizam a distribuição de previsão. O estudo da distribuição de previsão é de grande interesse na análise de séries temporais contínuas, especialmente na área de finanças (TAY; WALLIS, 2000).

Como proposto por Cardinal, Roy e Lambert (1999), Jung e Tremayne (2006), Kim e Park (2008) e Bisaglia e Gerolimetto (2019), métodos bootstrap podem ser utilizados para obter a distribuição de previsão h passos à frente. Assim, sendo $\mathbf{y} = (y_1, \dots, y_n)$ realizações de um processo ZI-INAR(1), o bootstrap para a previsão é dado pelas seguintes etapas:

1. Obtenha R_1 estimativas bootstrap $\{\hat{\theta}_r^*, r = 1, \dots, R_1\}$ utilizando um dos métodos apresentados na Seção 3.6.
2. Selecione aleatoriamente uma estimativa bootstrap do conjunto $\{\hat{\theta}_r^*, r = 1, \dots, R_1\}$ e gere $Y_{n+1}^*, \dots, Y_{n+h}^*$ recursivamente, utilizando a estimativa bootstrap selecionada, denotada por $\hat{\theta}^* = (\hat{\alpha}^*, \hat{\pi}^*, \hat{\lambda}^*)^\top$, como segue:

$$Y_{n+j}^* = \hat{\alpha}^* \circ Y_{n+j-1}^* + V_{n+j}^*, \quad j = 1, \dots, h,$$

em que $Y_n^* = y_n$, $V_{n+1}^*, \dots, V_{n+h}^*$ são v.a.'s iid seguindo distribuição ZI e $\hat{\alpha}^* \circ Y_{n+j-1}^* | Y_{n+j-1}^* = y_{n+j-1}^*$ segue uma distribuição Binomial com parâmetros y_{n+j-1}^* e $\hat{\alpha}^*$, se $y_{n+j-1}^* > 0$, e $\hat{\alpha}^* \circ Y_{n+j-1}^* | Y_{n+j-1}^* = y_{n+j-1}^*$ segue uma distribuição degenerada em zero, se $y_{n+j-1}^* = 0$, para $j = 1, \dots, h$.

3. Repita o Passo (2) R_2 vezes e utilize a distribuição empírica dos R_2 valores gerados para Y_{n+h}^* como a distribuição de previsão h passos à frente.

3.8 ESTUDOS DE SIMULAÇÃO

Nesta seção, apresentamos dois estudos de simulação. O primeiro tem como objetivo analisar a precisão das estimativas de máxima verossimilhança para os parâmetros do processo ZI-INAR(1). O segundo, tem como objetivo analisar as estimativas dos parâmetros do processo INAR(1) com inovações $\{U_t\}_{t \in \mathbb{Z}}$, i.e., não pertencentes à classe de distribuições ZI, quando a série é gerada por um processo INAR(1) com inovações pertencentes a classe das distribuições ZI.

Em ambos estudos de simulação, consideramos para as inovações as distribuições inflacionadas de zeros Poisson, Binomial Negativa e Poisson Inversa Gaussiana. Utilizamos o algoritmo EM para obter as estimativas de máxima verossimilhança dos parâmetros. Os valores iniciais adotados para os parâmetros no algoritmo foram os mesmos descritos na Seção 3.5 e o critério de convergência utilizado foi o critério baseado na aceleração de Aitken (MCLACHLAN; KRISHNAN, 2008).

3.8.1 Simulação 1: Propriedades assintóticas

Este estudo tem como objetivo avaliar a precisão das estimativas dos parâmetros para diferentes tamanhos amostrais. Foram consideradas amostras de tamanho $n = 100, 300, 500$ e 1000 . Para os parâmetros do processo, foram fixados $\alpha = 0.3$, $\pi = 0.3$ e 0.6 . Considerando as diferentes combinações de n e π , foram geradas $N = 300$ amostras dos seguintes processos ZI-INAR(1): ZIP-INAR(1) fixando $\lambda = 2$; ZINB-INAR(1) e ZIPIG-INAR(1) fixando $\mu = 2$, $\phi = 0.75, 1.5$ e 2.5 .

Após as $N = 300$ réplicas serem geradas para todos os cenários, foi analisada a precisão das estimativas dos parâmetros dos processos ZIP-INAR(1), ZINB-INAR(1) e ZIPIG-INAR(1) utilizando como medidas o viés relativo (VR) e a raiz quadrada do erro quadrático médio relativo (REQMR) dadas, respectivamente, por:

$$\text{VR}(\hat{\theta}_i) = \frac{1}{300} \sum_{j=1}^{300} \frac{\hat{\theta}_{ij} - \theta_i}{\theta_i},$$

$$\text{REQMR}(\hat{\theta}_i) = \sqrt{\frac{1}{300} \sum_{j=1}^{300} \left(\frac{\hat{\theta}_{ij} - \theta_i}{\theta_i} \right)^2},$$

em que $\hat{\theta}_{ij}$ é a estimativa de θ_i na j -ésima réplica e $\boldsymbol{\theta} = (\alpha, \pi, \boldsymbol{\lambda})$.

As Tabelas 2-4 apresentam os valores das medidas do VR e do REQMR das estimativas dos parâmetros. Nota-se que o VR das estimativas, em geral, estão em torno de zero, e o REQMR tende a diminuir quando aumentamos o valor de n . No entanto, comparando o desempenho das estimativas dos parâmetros para os processos ZINB-INAR(1) e ZIPIG-INAR(1), observamos que no geral, $\hat{\phi}$ apresentou pior desempenho, de acordo com as medidas analisadas. Ainda, é possível notar que as estimativas de α , μ e π apresentam valores menores do REQMR no processo ZIP-INAR(1) quando comparado com os processos ZINB-INAR(1) e ZIPIG-INAR(1), especialmente para valores de ϕ menores.

Sendo assim, concluímos por meio deste estudo de simulação, que as estimativas dos parâmetros se comportaram como o esperado, no sentido do viés das estimativas estarem em torno de zero e do erro quadrático médio diminuir para tamanhos maiores de amostra.

Tabela 2 – VR e REQMR das estimativas de máxima verossimilhança do processo ZIP-INAR(1) para diferentes tamanhos amostrais e diferentes valores de π , fixando $\alpha = 0.3$ e $\lambda = 2$.

π	n	$\hat{\alpha}$		$\hat{\pi}$		$\hat{\lambda}$	
		VR	REQMR	VR	REQMR	VR	REQMR
0.3	100	-0.0668	0.3141	-0.0764	0.3436	-0.0035	0.1163
	300	-0.0166	0.1635	-0.0090	0.1943	0.0058	0.0777
	500	-0.0229	0.1188	-0.0128	0.1372	0.0033	0.0579
	1000	-0.0020	0.0870	-0.0150	0.0995	-0.0024	0.0395
0.6	100	-0.0433	0.2362	-0.0212	0.1337	0.0007	0.1582
	300	-0.0118	0.1289	-0.0068	0.0707	-0.0058	0.0895
	500	-0.0117	0.0985	-0.0050	0.0488	0.0021	0.0650
	1000	-0.0095	0.0750	-0.0031	0.0388	-0.0040	0.0472

Tabela 3 – VR e REQMR das estimativas de máxima verossimilhança do processo ZINB-INAR(1) para diferentes tamanhos amostrais e diferentes valores de π , fixando $\alpha = 0.3$, $\mu = 2$ e $\phi = 0.75, 1.5$ e 2.5

ϕ	π	n	$\hat{\alpha}$		$\hat{\pi}$		$\hat{\mu}$		$\hat{\phi}$	
			VR	REQMR	VR	REQMR	VR	REQMR	VR	REQMR
0.75	0.3	100	0.0051	0.2080	-0.0001	0.7010	0.0971	0.3614	0.3417	1.8875
		300	0.0042	0.1117	-0.0365	0.5423	0.0344	0.2355	0.1165	0.6861
		500	0.0022	0.0842	-0.0238	0.4571	0.0143	0.1873	0.0696	0.4845
		1000	0.0019	0.0595	-0.0424	0.3500	-0.0048	0.1376	0.0111	0.2943
	0.6	100	0.0052	0.1893	-0.1443	0.4576	0.0306	0.4736	0.2758	3.2307
		300	-0.0148	0.1091	-0.0688	0.2908	0.0106	0.2980	0.0990	0.9289
		500	0.0083	0.0830	-0.0684	0.2552	-0.0158	0.2584	0.0125	0.6428
		1000	-0.0021	0.0612	-0.0242	0.1472	-0.0076	0.1692	0.0602	0.3940
1.5	0.3	100	-0.0157	0.2241	-0.0666	0.6120	0.0183	0.2399	0.2144	1.4996
		300	-0.0015	0.1219	-0.0834	0.4179	-0.0147	0.1604	0.0298	0.6910
		500	-0.0019	0.0899	-0.0185	0.3245	0.0084	0.1294	0.0585	0.4348
		1000	0.0020	0.0687	-0.0333	0.2436	-0.0043	0.0950	-0.0039	0.2898
	0.6	100	-0.0200	0.2105	-0.1236	0.3800	-0.0118	0.3557	0.0959	2.1109
		300	0.0005	0.1204	-0.0407	0.1930	-0.0130	0.2056	0.0401	1.1545
		500	-0.0109	0.0949	-0.0256	0.1351	-0.0167	0.1607	0.0213	0.6549
		1000	-0.0004	0.0603	-0.0218	0.0883	-0.0184	0.1161	-0.0314	0.3832
2.5	0.3	100	-0.0388	0.2513	-0.1162	0.5348	0.0030	0.2104	0.0319	0.9771
		300	-0.0050	0.1370	-0.0642	0.3803	-0.0077	0.1390	0.1027	0.9715
		500	-0.0212	0.1089	-0.0502	0.2699	-0.0061	0.1037	0.0178	0.5092
		1000	-0.0066	0.0764	-0.0281	0.1829	-0.0082	0.0755	-0.0169	0.3569
	0.6	100	-0.0169	0.2278	-0.1465	0.3334	-0.0959	0.2975	-0.1273	1.3087
		300	-0.0125	0.1194	-0.0165	0.1262	-0.0176	0.1730	0.0525	1.0700
		500	-0.0030	0.0889	-0.0088	0.0954	-0.0022	0.1286	0.0730	0.8590
		1000	-0.0006	0.0697	-0.0065	0.0690	-0.0025	0.0977	-0.0000	0.4792

Tabela 4 – VR e REQMR das estimativas de máxima verossimilhança do processo ZIPIG-INAR(1) para diferentes tamanhos amostrais e diferentes valores de π , fixando $\alpha = 0.3$, $\mu = 2$ e $\phi = 0.75, 1.5$ e 2.5

ϕ	π	n	$\hat{\alpha}$		$\hat{\pi}$		$\hat{\mu}$		$\hat{\phi}$	
			VR	REQMR	VR	REQMR	VR	REQMR	VR	REQMR
0.75	0.3	100	0.0010	0.2044	-0.0046	0.5836	0.0391	0.2632	0.1688	1.8383
		300	-0.0056	0.1175	-0.0665	0.3940	-0.0078	0.1739	0.0431	0.5878
		500	-0.0002	0.0887	-0.0243	0.2975	0.0004	0.1293	0.0237	0.3576
		1000	-0.0016	0.0614	-0.0117	0.2135	-0.0004	0.0918	0.0301	0.2578
	0.6	100	-0.0205	0.2034	-0.0825	0.3159	0.0369	0.3589	0.2246	2.6273
		300	-0.0007	0.1125	-0.0345	0.1567	-0.0168	0.2096	-0.0025	0.7321
		500	-0.0073	0.0871	-0.0121	0.1040	-0.0096	0.1539	0.0717	0.5155
		1000	-0.0027	0.0677	-0.0095	0.0790	-0.0110	0.1204	0.0187	0.3109
1.5	0.3	100	-0.0340	0.2193	-0.0740	0.5251	0.0183	0.2156	0.1622	1.3835
		300	-0.0068	0.1367	-0.0243	0.3257	0.0016	0.1346	0.0398	0.7542
		500	-0.0050	0.1032	-0.0162	0.2326	0.0007	0.0959	0.0268	0.4768
		1000	-0.0065	0.0702	-0.0273	0.1705	-0.0001	0.0762	-0.0163	0.2293
	0.6	100	-0.0320	0.2046	-0.0452	0.2319	0.0160	0.2772	0.1455	1.6193
		300	-0.0052	0.1066	-0.0212	0.1103	-0.0034	0.1659	0.0623	0.9166
		500	-0.0051	0.0913	-0.0096	0.0885	-0.0015	0.1253	0.0543	0.9582
		1000	0.0022	0.0640	0.0025	0.0573	0.0051	0.0942	0.0481	0.3459
2.5	0.3	100	-0.0437	0.2472	-0.0750	0.4556	0.0187	0.1978	0.1961	1.3687
		300	-0.0095	0.1382	-0.0194	0.2708	-0.0006	0.1141	0.0972	0.9488
		500	-0.0040	0.1188	0.0002	0.2095	0.0061	0.0938	0.0628	0.4610
		1000	-0.0042	0.0720	-0.0017	0.1384	0.0070	0.0621	0.0338	0.3133
	0.6	100	-0.0486	0.2160	-0.0850	0.2530	-0.0336	0.2658	-0.0715	1.1974
		300	-0.0107	0.1299	-0.0161	0.0976	-0.0046	0.1420	0.0965	0.9962
		500	-0.0088	0.0918	-0.0143	0.0777	-0.0097	0.1051	0.0423	0.7257
		1000	0.0025	0.0653	-0.0031	0.0503	-0.0027	0.0776	0.0077	0.3785

3.8.2 Simulação 2: Efeito da má especificação da distribuição das inovações

Este estudo de simulação tem como objetivo avaliar o impacto nas estimativas dos parâmetros de um processo INAR(1) com inovações seguindo distribuições ZI, quando especificamos para tais inovações uma distribuição não pertencente à classe das distribuições ZI. Para realizar o estudo, geramos apenas processos com inovações seguindo distribuições ZI, e para cada processo gerado, estimamos os parâmetros ajustando dois processos: INAR(1) com inovações seguindo distribuições ZI (especificação correta) e INAR(1) com inovações não pertencente à classe das distribuições ZI (especificação incorreta). Em particular, geramos os processos:

- ZIP-INAR(1), e estimamos os parâmetros ajustando os processos Po-INAR(1) e ZIP-INAR(1);
- ZINB-INAR(1), e estimamos os parâmetros ajustando os processos NB-INAR(1) e ZINB-INAR(1);
- ZIPIG-INAR(1), e estimamos os parâmetros ajustando os processos PIG-INAR(1) e ZIPIG-INAR(1);

Para gerar os processos ZIP-INAR(1), ZINB-INAR(1) e ZIPIG-INAR(1), utilizamos amostras de tamanho $n = 300$ e fixamos os parâmetros em $\alpha = 0.3$ e $\pi = 0.1, 0.3, 0.5$ e 0.7 (a ideia é analisar se o impacto nas estimativas se alteram para diferentes valores de π) para os três casos. Considerando os diferentes valores de π , utilizamos $N = 300$ réplicas dos processos ZIP-INAR(1), fixando $\lambda = 2$ e, ZINB-INAR(1) e ZIPIG-INAR(1), fixando $\mu = 2$ e $\phi = 0.75, 1.5$, e 2.5 .

Nas Tabelas 5-7 (assim como nas Figuras 7-9 apresentadas no Apêndice A), nota-se nos três processos INAR(1) com inovações $\{U_t\}_{t \in \mathbb{Z}}$, que o parâmetro cujas estimativas apresentaram maior viés foi o parâmetro μ . As estimativas deste parâmetro subestimam μ e o viés tende a aumentar para valores de π maiores. Observamos que $\hat{\alpha}$ também subestima o verdadeiro valor do parâmetro nos três processos INAR(1) com inovações $\{U_t\}_{t \in \mathbb{Z}}$. No entanto, nos processos NB-INAR(1) e PIG-INAR(1), a estimativa deste parâmetro aparentemente não é afetada para os diferentes valores de π considerados, enquanto que no processo Po-INAR(1), $\hat{\alpha}$ apresenta maior viés para valores de π a partir de 0.3 . Com relação as estimativas de ϕ nos processos NB-INAR(1) e PIG-INAR(1), notamos o mesmo problema observado em $\hat{\mu}$, ou seja, $\hat{\phi}$ subestima ϕ , elevando seu viés quando π aumenta.

Sendo assim, concluímos que especificar incorretamente as distribuições Poisson, NB e PIG para as inovações, quando tais inovações seguem distribuições ZIP, ZINB, e ZIPIG, respectivamente, afeta as estimativas dos parâmetros, em geral.

Tabela 5 – VR das estimativas de máxima verossimilhança dos processos Po-INAR(1) e ZIP-INAR(1) em amostras de tamanho $n = 300$ para diferentes valores de π , fixando $\alpha = 0.3$ e $\lambda = 2$.

π	Processo	VR	
		$\hat{\alpha}$	$\hat{\lambda}$
0.1	Po-INAR(1)	-0.0804	-0.0680
	ZIP-INAR(1)	-0.0347	0.0094
0.3	Po-INAR(1)	-0.2015	-0.2392
	ZIP-INAR(1)	-0.0202	0.0058
0.5	Po-INAR(1)	-0.2233	-0.4544
	ZIP-INAR(1)	-0.0082	-0.0013
0.7	Po-INAR(1)	-0.2235	-0.6713
	ZIP-INAR(1)	-0.0089	0.0055

Tabela 6 – VR das estimativas de máxima verossimilhança dos processos NB-INAR(1) e ZINB-INAR(1) em amostras de tamanho $n = 300$ para diferentes valores de π e ϕ , fixando $\alpha = 0.3$, $\mu = 2$.

ϕ	π	Processo	VR		
			$\hat{\alpha}$	$\hat{\mu}$	$\hat{\phi}$
0.75	0.1	NB-INAR(1)	-0.0191	-0.0904	-0.1682
		ZINB-INAR(1)	0.0009	0.0636	0.1463
	0.3	NB-INAR(1)	-0.0284	-0.2877	-0.4698
		ZINB-INAR(1)	-0.0064	0.0433	0.0895
	0.5	NB-INAR(1)	-0.0100	-0.4940	-0.6903
		ZINB-INAR(1)	0.0058	-0.0328	-0.0491
	0.7	NB-INAR(1)	-0.0184	-0.7040	-0.8429
		ZINB-INAR(1)	-0.0089	-0.0351	0.0640
	1.5	NB-INAR(1)	-0.0061	-0.0992	-0.2432
		ZINB-INAR(1)	0.0208	0.0174	0.0209
2.5	0.3	NB-INAR(1)	-0.0348	-0.2887	-0.5525
		ZINB-INAR(1)	-0.0008	-0.0087	0.0388
	0.5	NB-INAR(1)	-0.0309	-0.4921	-0.7569
		ZINB-INAR(1)	-0.0028	0.0070	0.0679
	0.7	NB-INAR(1)	-0.0159	-0.6970	-0.8899
		ZINB-INAR(1)	-0.0016	-0.0170	0.0245
	0.1	NB-INAR(1)	-0.0439	-0.0805	-0.2559
		ZINB-INAR(1)	-0.0167	0.0164	0.0380
	0.3	NB-INAR(1)	-0.0569	-0.2812	-0.6252
		ZINB-INAR(1)	-0.0116	0.0038	0.1484
	0.5	NB-INAR(1)	-0.0386	-0.4878	-0.8142
		ZINB-INAR(1)	-0.0060	-0.0089	0.0548
	0.7	NB-INAR(1)	-0.0191	-0.6974	-0.9181
		ZINB-INAR(1)	-0.0010	-0.0182	0.0452

Tabela 7 – VR das estimativas de máxima verossimilhança dos processos PIG-INAR(1) e ZIPIG-INAR(1) em amostras de tamanho $n = 300$ para diferentes valores de π e ϕ , fixando $\alpha = 0.3$, $\mu = 2$.

ϕ	π	Processo	VR		
			$\hat{\alpha}$	$\hat{\mu}$	$\hat{\phi}$
0.75	0.1	PIG-INAR(1)	-0.0478	-0.0878	-0.1590
		ZIPIG-INAR(1)	-0.0169	0.0185	0.0526
	0.3	PIG-INAR(1)	-0.0521	-0.2872	-0.4945
		ZIPIG-INAR(1)	-0.0101	0.0216	0.0574
	0.5	PIG-INAR(1)	-0.0462	-0.4948	-0.7215
		ZIPIG-INAR(1)	-0.0084	-0.0064	0.0268
	0.7	PIG-INAR(1)	-0.0307	-0.6939	-0.8758
		ZIPIG-INAR(1)	-0.0081	0.0424	0.1952
1.5	0.1	PIG-INAR(1)	-0.0572	-0.0781	-0.1764
		ZIPIG-INAR(1)	-0.0099	0.0126	-0.0018
	0.3	PIG-INAR(1)	-0.0751	-0.2775	-0.5605
		ZIPIG-INAR(1)	-0.0083	0.0042	0.0262
	0.5	PIG-INAR(1)	-0.0616	-0.4906	-0.7857
		ZIPIG-INAR(1)	-0.0059	0.0023	0.0068
	0.7	PIG-INAR(1)	-0.0466	-0.6965	-0.9165
		ZIPIG-INAR(1)	-0.0158	0.0205	0.2048
2.5	0.1	PIG-INAR(1)	-0.0586	-0.0804	-0.2482
		ZIPIG-INAR(1)	-0.0142	0.0193	0.0916
	0.3	PIG-INAR(1)	-0.0899	-0.2693	-0.6166
		ZIPIG-INAR(1)	-0.0093	0.0022	0.0415
	0.5	PIG-INAR(1)	-0.0786	-0.4771	-0.8279
		ZIPIG-INAR(1)	-0.0057	0.0160	0.1043
	0.7	PIG-INAR(1)	-0.0444	-0.6989	-0.9371
		ZIPIG-INAR(1)	-0.0081	-0.0067	0.0252

3.9 APLICAÇÃO

Nesta seção, será apresentada uma aplicação a dados reais utilizando os processos ZI-INAR(1) e a metodologia proposta. O conjunto de dados é referente a números mensais de crimes relacionado a drogas no distrito 2206 de São Petersburgo, no período de janeiro de 1990 até dezembro de 2001, disponível no site (<http://www.forecastingprinciples.com>). Os crimes são contabilizados pelo número de relatórios mensais elaborados pela polícia com relação ao tipo de crime em questão.

Na Tabela 8 são apresentadas algumas estatísticas descritivas da série, em que é possível observar uma porcentagem de 43.1% de zeros, e média e variância amostral iguais a 2.111 e 12.911, respectivamente. Pela análise destes resultados, aparentemente um processo ZI-INAR(1) é adequado.

A Figura 1 apresenta o gráfico da série, o gráfico de barras, e a função de autocorrelação parcial (FACP). É possível observar pelo gráfico da série dois valores maiores do que o esperado, representados pelas observações $y_{58} = 19$ e $y_{59} = 29$, que ocorreram nos meses de outubro e novembro de 1994. O gráfico da FACP sugere que um modelo autorregressivo de primeira ordem seja adequado, uma vez que apenas a autocorrelação de

$lag\ 1$ é significativa.

Tabela 8 – Estatísticas descritivas da série *número de crimes no distrito 2206*.

n	144
Mínimo	0
Média	2.111
Mediana	1
Máximo	29
Variância	12.911
Proporção de zeros	0.431

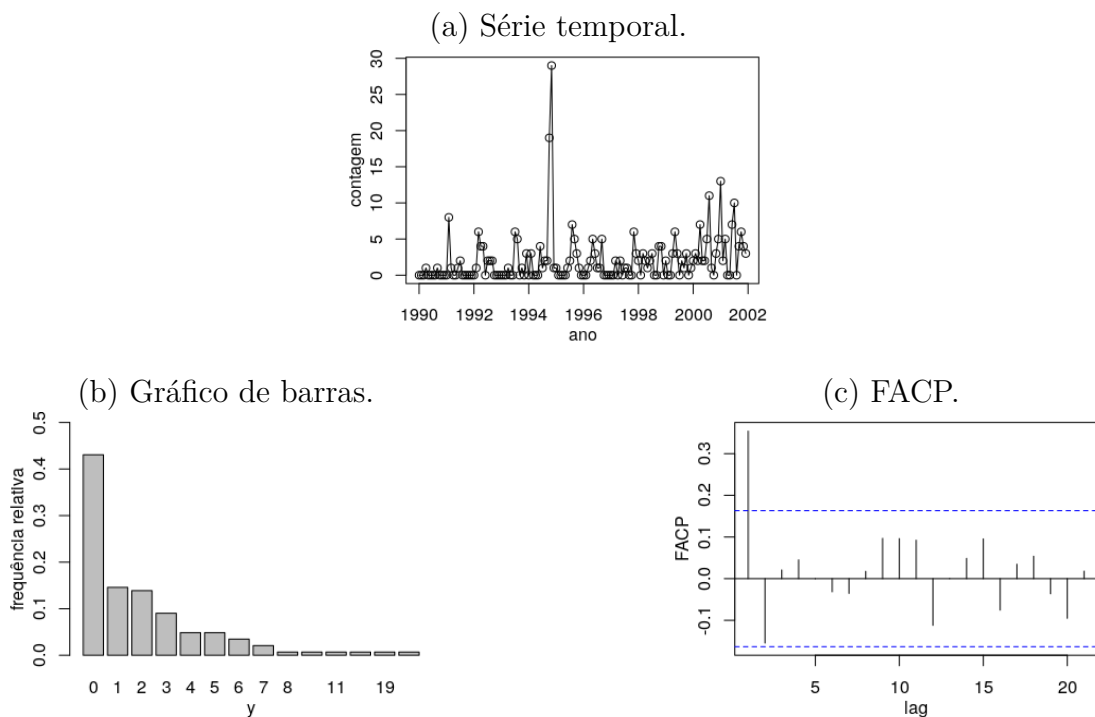


Figura 1 – Série *número de crimes no distrito 2206*.

Estimação:

Na Tabela 9 são apresentadas as estimativas dos parâmetros e o critério de informação de Akaike (AIC). As estimativas dos parâmetros foram obtidas utilizando a metodologia apresentada na Seção 3.5 e a medida do AIC foi obtida utilizando a função de log-verossimilhança condicional dada em (3.18). Por meio da análise desta medida, sugerimos que o processo ZIPIG-INAR(1) é mais adequado para modelar a série.

A Tabela 10 apresenta, para cada processo, a estimativa da probabilidade de zero, definida como $\hat{P}_0(Y_t = 0)$, assim como a estimativa do valor esperado da v.a. N : *número de zeros consecutivos até o primeiro valor diferente de zero*, definida como

Tabela 9 – Estimativa dos parâmetros para os processos INAR(1) com inovações Poisson, ZIP, NB, ZINB, PIG e ZIPIG. Série: *número de crimes no distrito 2206*.

Processo	$\hat{\alpha}$	$\hat{\pi}$	$\hat{\mu}$	$\hat{\phi}$	AIC
Po-INAR(1)	0.212	-	1.679	-	590.74
ZIP-INAR(1)	0.181	0.512	3.577	-	565.77
NB-INAR(1)	0.071	-	1.977	0.471	550.43
ZINB-INAR(1)	0.070	0.138	2.296	0.630	552.20
PIG-INAR(1)	0.072	-	1.973	0.336	554.53
ZIPIG-INAR(1)	0.065	0.325	2.946	0.903	549.41

$\hat{E}_0(N)$. Note que o processo Po-INAR(1) não é adequado, uma vez que as estimativas da probabilidade de zero (0.1187) e do valor esperado da v.a. N (1.2293) são inferiores à proporção de zeros e à média do número de zeros consecutivos observados na série (0.4306 e 2.2963, respectivamente). Note ainda que os processos com inovações ZI, i.e., ZIP-INAR(1), ZINB-INAR(1) e ZIPIG-INAR(1), estimaram melhor a probabilidade de zero e a esperança da v.a. N (no sentido de se aproximarem mais das quantidades empíricas), quando comparado com os processos Po-INAR(1), NB-INAR(1) e PIG-INAR(1), respectivamente.

Tabela 10 – Estimativa da probabilidade de zero e do valor esperado do *número de zeros consecutivos até o primeiro valor diferente de zero*.

Processo	$\hat{P}_0(Y_t = 0)$	$\hat{E}_0(N)$
Po-INAR(1)	0.1187	1.2293
ZIP-INAR(1)	0.3769	2.5213
NB-INAR(1)	0.4028	1.8528
ZINB-INAR(1)	0.4080	1.8719
PIG-INAR(1)	0.3697	1.7295
ZIPIG-INAR(1)	0.4110	1.8679
Empírico	0.4306	2.2963

A Figura 2 apresenta as distribuições bootstrap dos estimadores dos parâmetros do processo ZIPIG-INAR(1), obtidas utilizando os métodos de bootstrap paramétrico e bootstrap circular em blocos (veja Seção 3.6). Como sugerido por Bisaglia e Gerolimetto (2019), para o método por reamostragem em blocos, consideramos tamanhos de blocos $l = \sqrt{n}$, i.e., $l = 12$. Utilizamos $R = 1000$ réplicas de bootstrap para os dois métodos propostos. É possível notar que as distribuições bootstrap obtidas pelo método paramétrico apresentam comportamentos similares às obtidas pelo método por reamostragem em blocos circulares, com exceção da distribuição de $\hat{\phi}$. Neste caso, no bootstrap circular em blocos, observamos algumas estimativas de ϕ distantes da média (1.385), como por exemplo, $\hat{\phi}_{205} = 19.98$, $\hat{\phi}_{284} = 22.71$, $\hat{\phi}_{598} = 17.70$ e $\hat{\phi}_{646} = 10.10$.

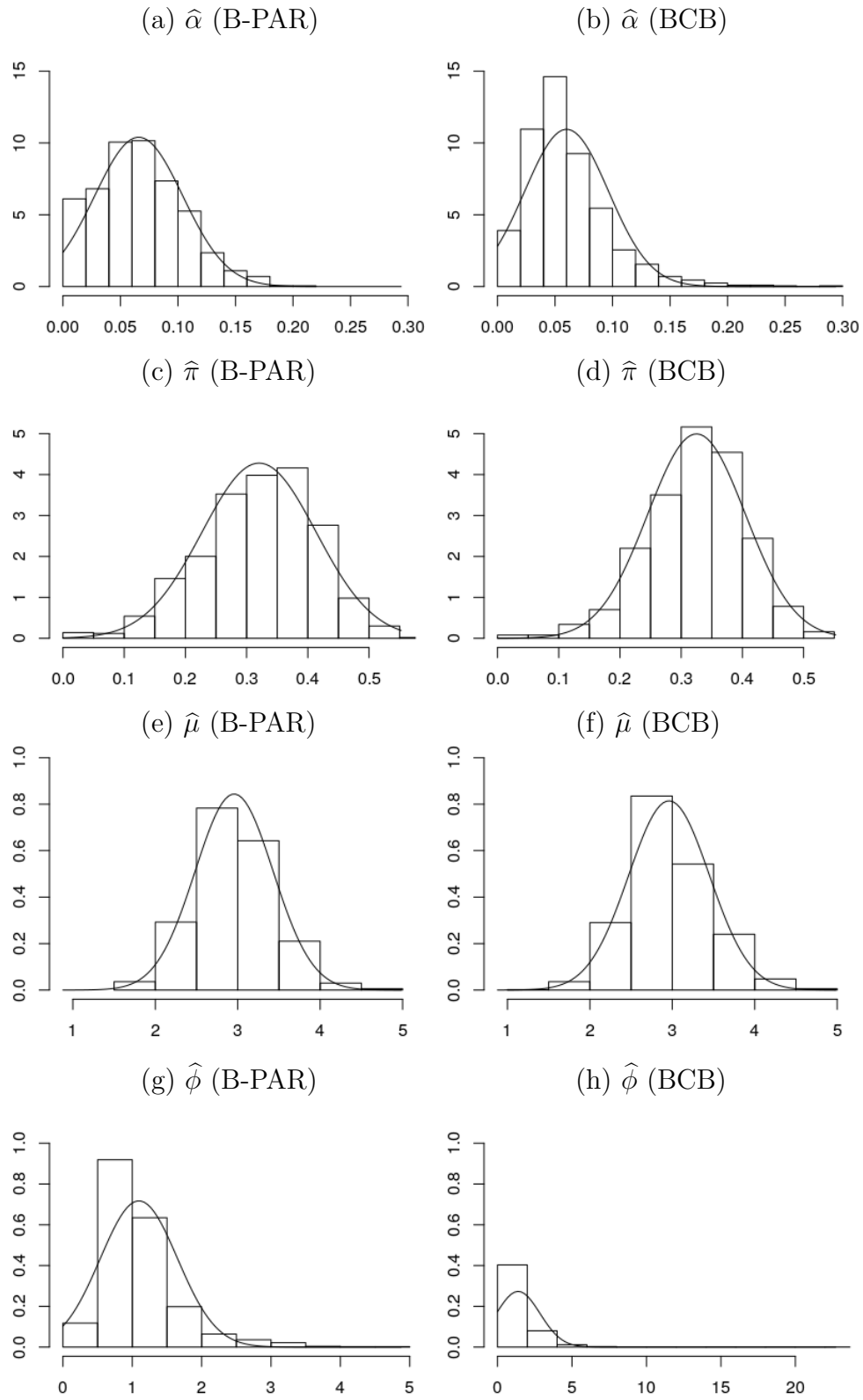


Figura 2 – Distribuição dos estimadores dos parâmetros do processo ZIPIG-INAR(1) obtidas pelos métodos de bootstrap paramétrico (B-PAR) e bootstrap circular em blocos (BCB).

A Tabela 11 apresenta as estimativas e intervalos bootstrap baseados na variância assintótica, para os parâmetros do processo ZIPIG-INAR(1). Conforme observamos, as estimativas e os intervalos de confiança para os parâmetros α , π e μ , obtidos pelos dois métodos de bootstrap, apresentaram valores similares. No entanto, o mesmo padrão não pode ser observado quando comparamos tais estimativas e intervalos de confiança para o parâmetro ϕ . Relacionado aos intervalos bootstrap para o parâmetro π , note que estes podem ser utilizados para testar $H_0 : \pi = 0$ *versus* $H_1 : \pi > 0$, isto é, testar H_0 : as inovações seguem distribuição PIG *versus* H_1 : as inovações seguem distribuição ZIPIG. Uma vez que os intervalos não contém o valor zero, assim como na análise do AIC, sugerimos utilizar o processo ZIPIG-INAR(1) quando comparado com o processo PIG-INAR(1).

Tabela 11 – Estimativas e intervalos de confiança para os parâmetros do processo ZIPIG-INAR(1). B-PAR (Bootstrap paramétrico). BCB (Bootstrap circular em blocos).

Parâmetro	Estimativa			Desv. Pad.		Intervalo bootstrap	
	EM	B-PAR	BCB	B-PAR	BCB	B-PAR	BCB
α	0.065	0.066	0.060	0.038	0.036	[0,0.141]	[0,0.131]
π	0.325	0.320	0.325	0.093	0.080	[0.138,0.503]	[0.169,0.482]
μ	2.947	2.956	2.960	0.473	0.490	[2.029,3.883]	[2.000,3.921]
ϕ	0.903	1.095	1.385	0.556	1.459	[0.005,2.185]	(0,4.245]

Previsão:

A Tabela 12 apresenta o Erro Absoluto Médio de Previsão 1 Passo à Frente (EAMP(1)), dado na Equação (3.40), com o intuito de comparar as previsões dos diferentes processos ajustados. Para o cálculo da medida do EAMP(1), utilizamos as últimas 12 observações da série para a previsão, enquanto que as 132 primeiras observações foram utilizadas para a estimação dos parâmetros, ou seja, utilizamos $m = 12$. Nota-se que os processos cujas inovações seguem distribuições ZIP, ZINB e ZIPIG apresentaram melhores resultados quando comparado com processos com inovações seguindo distribuições Poisson, Binomial Negativa e Poisson Inversa Gaussiana, respectivamente. Na comparação de todos os processos, aquele que apresentou menor EAMP foi o processo ZIPIG-INAR(1).

Tabela 12 – Erro Absoluto Médio de Previsão 1 Passo à Frente (EAMP(1)) para a série *número de crimes no distrito 2206*.

Processo	EAMP
Po-INAR(1)	3.6578
ZIP-INAR(1)	3.6505
NB-INAR(1)	3.5542
ZINB-INAR(1)	3.5541
PIG-INAR(1)	3.5572
ZIPIG-INAR(1)	3.5480

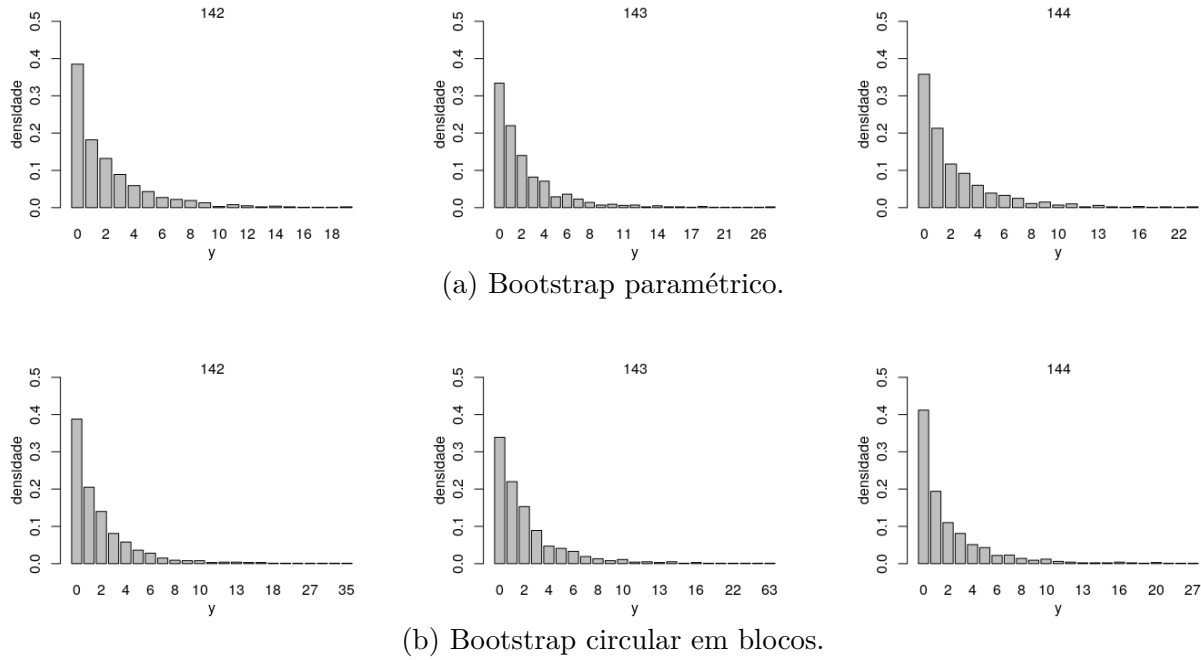


Figura 3 – Distribuição de previsão $h = 1, 2$ e 3 passos à frente para a série temporal *número de crimes no distrito 2206*.

A Figura 3 apresenta as distribuições de previsão $h = 1, 2$ e 3 passos à frente das três últimas observações da série (y_{142} , y_{143} e y_{144}) obtidas pelo ajuste do processo ZIPIG-INAR(1), i.e., a distribuição de $Y_{142}|Y_{141} = y_{141}$, $Y_{143}|Y_{141} = y_{141}$ e $Y_{144}|Y_{141} = y_{141}$. Para obter estas distribuições de previsão seguimos os passos (1)-(3) dados na Seção 3.7. No passo (1), geramos $R_1 = 100$ estimativas bootstrap, utilizando as observações $y_1, \dots, y_{141}$ da série, e nos passos (2) e (3), geramos $R_2 = 1000$ valores para y_{142}^* , y_{143}^* e y_{144}^* . É possível perceber que as distribuições de previsão obtidas pelo método de bootstrap paramétrico apresentam comportamentos similares às obtidas pelo método de bootstrap em blocos circulares. Note que em todos casos, observamos um comportamento decrescente destas distribuições, tendo como valor com maior probabilidade, o zero. Utilizando o método de bootstrap paramétrico, estima-se que as probabilidades de observar os três últimos valores da série ($y_{142} = 6$, $y_{143} = 4$ e $y_{144} = 3$), dado $y_1, \dots, y_{141}$, são de 0,027, 0,071

e 0,092, respectivamente, e utilizando o método de reamostragem em blocos circulares, estas probabilidades são de 0,028, 0,047 e 0,081.

4 PROCESSOS ZI-INAR(p)

Na literatura não existem muitos trabalhos relacionados a séries temporais de contagem com inflação de zeros e de ordem $p > 1$. Sendo assim, neste capítulo propomos o processo autorregressivo de valores inteiros não negativos de ordem p com inovação seguindo distribuições ZI, ZI-INAR(p), uma extensão do processo ZI-INAR(1).

Neste capítulo, apresentamos a definição do processo INAR(p), proposto por Du e Li (1991). Em seguida, definimos o novo processo ZI-INAR(p), apresentamos propriedades do processo, métodos de estimação dos parâmetros (por máxima verossimilhança via algoritmo EM) e métodos de previsão. Realizamos um estudo de simulação e uma aplicação a dados reais para ilustrar a metodologia proposta.

4.1 PROCESSO INAR(p)

O processo autorregressivo de valores inteiros não negativos de ordem p é uma extensão do processo INAR(1), denotado por INAR(p), e definido como:

$$Y_t = \sum_{i=1}^p \alpha_i \circ Y_{t-i} + V_t, \quad t \in \mathbb{Z}, \quad (4.1)$$

em que, dado $Y_{t-1} = y_{t-1}, \dots, Y_{t-p} = y_{t-p}$, as v.a's $\alpha_i \circ Y_{t-i}$, com $\alpha_i \in [0, 1]$, $i = 1, \dots, p$, são condicionalmente independentes e a sequência de inovações $\{V_t\}_{t \in \mathbb{Z}}$ são v.a's iid com média μ_V e variância finita σ_V^2 , independentes de $Y_{t-1}, Y_{t-2}, \dots$.

As condições de estacionariedade de um processo INAR(p) são que as raízes do polinômio

$$z^p - \alpha_1 z^{p-1} - \dots - \alpha_{p-1} z - \alpha_p$$

estejam dentro do círculo unitário, isto é, $|z| < 1$ (DU; LI, 1991). Assim, Dion, Gauthier e Latour (1995) mostraram que esta condição é equivalente a

$$\sum_{i=1}^p \alpha_i < 1.$$

Sob condições de estacionariedade, a esperança e a variância de Y_t , $\forall t \in \mathbb{Z}$, são dadas por:

$$E(Y_t) = \frac{\mu_V}{1 - \sum_{i=1}^p \alpha_i} \quad \text{e} \quad Var(Y_t) = \frac{E(Y_t) \sum_{i=1}^p \alpha_i (1 - \alpha_i) + \sigma_V^2}{1 - \sum_{i=1}^p \alpha_i \rho(i)}, \quad (4.2)$$

em que μ_V e σ_V^2 representam a média e a variância das inovações V_t , respectivamente, e $\rho(i)$ é a autocorrelação de defasagem i .

4.2 DEFINIÇÃO E PROPRIEDADES

Definição 4.1. Dizemos que o processo $\{Y_t\}_{t \in \mathbb{Z}}$ definido pela Equação (4.1) é um processo autorregressivo para valores inteiros de ordem p , com inovações seguindo distribuição inflacionada de zeros, denotado por $ZI-INAR(p)$, se a v.a. $V_t \sim ZI(\pi, \boldsymbol{\lambda}, h_{U_t})$.

A seguir, apresentamos algumas propriedades do processo $ZI-INAR(p)$, considerando os casos particulares, em que as inovações seguem distribuição ZIP, ZINB, ou ZIPIG.

Proposição 4.1. Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo $ZI-INAR(p)$ estacionário, então a esperança e a variância de Y_t são dadas por:

$$E(Y_t) = \frac{(1 - \pi)E(U_t)}{1 - \sum_{i=1}^p \alpha_i}, \quad (4.3)$$

e

$$Var(Y_t) = \frac{(1 - \pi) \left[E(U_t) \sum_{i=1}^p \alpha_i (1 - \alpha_i) + \left(1 - \sum_{i=1}^p \alpha_i \right) (Var(U_t) + \pi E^2(U_t)) \right]}{\left(1 - \sum_{i=1}^p \alpha_i \rho(i) \right) \left(1 - \sum_{i=1}^p \alpha_i \right)} \quad (4.4)$$

Prova. Utilizando a esperança do processo $INAR(p)$ dada na Equação (4.2) e a esperança do modelo ZI , segue de imediato que a Equação (4.3) é válida. Por outro lado, a variância de Y_t é obtida por:

$$\begin{aligned} Var(Y_t) &= \frac{E(Y_t) \sum_{i=1}^p \alpha_i (1 - \alpha_i) + \sigma_V^2}{1 - \sum_{i=1}^p \alpha_i \rho(i)} \\ &= \frac{(1 - \pi) \left[E(U_t) \sum_{i=1}^p \alpha_i (1 - \alpha_i) + \left(1 - \sum_{i=1}^p \alpha_i \right) (Var(U_t) + \pi E^2(U_t)) \right]}{\left(1 - \sum_{i=1}^p \alpha_i \rho(i) \right) \left(1 - \sum_{i=1}^p \alpha_i \right)}, \quad (4.5) \end{aligned}$$

em que utilizamos a variância do modelo ZI para obter (4.5). $\square$

Assim, a razão entre a variância e a média é dada por:

$$\frac{Var(Y_t)}{E(Y_t)} = \frac{E(U_t) \sum_{i=1}^p \alpha_i (1 - \alpha_i) + \left(1 - \sum_{i=1}^p \alpha_i \right) (Var(U_t) + \pi E^2(U_t))}{E(U_t) \left(1 - \sum_{i=1}^p \alpha_i \rho(i) \right)}$$

Em particular, temos que:

- Se $V_t \sim \text{ZIP}(\pi, \lambda)$, então $U_t \sim \text{Poisson}(\lambda)$. Logo:

$$E(Y_t) = \frac{(1 - \pi)\lambda}{1 - \sum_{i=1}^p \alpha_i}$$

e

$$\begin{aligned} \text{Var}(Y_t) &= \frac{(1 - \pi)\lambda \left[\sum_{i=1}^p \alpha_i(1 - \alpha_i) + \left(1 - \sum_{i=1}^p \alpha_i\right)(1 + \pi\lambda) \right]}{\left(1 - \sum_{i=1}^p \alpha_i \rho(i)\right) \left(1 - \sum_{i=1}^p \alpha_i\right)} \\ &= \frac{E(Y_t) \left[1 - \sum_{i=1}^p \alpha_i^2 + \pi\lambda \left(1 - \sum_{i=1}^p \alpha_i\right) \right]}{1 - \sum_{i=1}^p \alpha_i \rho(i)}. \end{aligned}$$

Neste caso, a razão entre a variância e a média é dada por:

$$\frac{\text{Var}(Y_t)}{E(Y_t)} = \frac{1 - \sum_{i=1}^p \alpha_i^2 + \pi\lambda \left(1 - \sum_{i=1}^p \alpha_i\right)}{1 - \sum_{i=1}^p \alpha_i \rho(i)}.$$

- Para os casos em que $V_t \sim \text{ZINB}(\pi, \mu, \phi)$ e $V_t \sim \text{ZIPIG}(\pi, \mu, \phi)$, temos que $E(U_t) = \mu$ e $\text{Var}(U_t) = \mu + \mu^2/\phi$. Portanto:

$$E(Y_t) = \frac{(1 - \pi)\mu}{1 - \sum_{i=1}^p \alpha_i}, \quad (4.6)$$

e

$$\begin{aligned} \text{Var}(Y_t) &= \frac{(1 - \pi)\mu \left[\sum_{i=1}^p \alpha_i(1 - \alpha_i) + \left(1 - \sum_{i=1}^p \alpha_i\right)(1 + \pi\mu + \mu/\phi) \right]}{\left(1 - \sum_{i=1}^p \alpha_i \rho(i)\right) \left(1 - \sum_{i=1}^p \alpha_i\right)} \\ &= \frac{E(Y_t) \left[1 - \sum_{i=1}^p \alpha_i^2 + (\pi\mu + \mu/\phi) \left(1 - \sum_{i=1}^p \alpha_i\right) \right]}{1 - \sum_{i=1}^p \alpha_i \rho(i)}. \end{aligned} \quad (4.7)$$

Assim, a razão entre a variância e a média do processo pode ser expressa por:

$$\frac{\text{Var}(Y_t)}{E(Y_t)} = \frac{1 - \sum_{i=1}^p \alpha_i^2 + (1 + \pi\mu + \mu/\phi) \left(1 - \sum_{i=1}^p \alpha_i\right)}{1 - \sum_{i=1}^p \alpha_i \rho(i)} \quad (4.8)$$

4.3 DISTRIBUIÇÃO CONDICIONAL

Nesta seção, mostramos como obter uma expressão para a probabilidade condicional do processo ZI-INAR(p), com o intuito de utiliza-la no cálculo da função de verossimilhança. A estratégia que utilizamos é a mesma proposta por Bu, McCabe e Kaddour (2008). Assim, sendo $\{Y_t\}_{t \in \mathbb{Z}}$ um processo ZI-INAR(p), e as v.a.'s $\mathbf{Y}_{t(-p)} = (Y_{t-1}, \dots, Y_{t-p})$ e $S_{t,i} = \alpha_i \circ Y_{t-i}$, para $i = 1, \dots, p$. A ideia é considerar Y_t como a convolução das v.a.'s $S_{t,1} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$ e $\sum_{i=2}^p S_{t,i} + V_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$, que por definição são v.a.'s independentes. Em seguida, consideramos $\sum_{i=2}^p S_{t,i} + V_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$ como a convolução das v.a.'s $S_{t,2} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$ e $\sum_{i=3}^p S_{t,i} + V_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$, e assim por diante. Assim, a probabilidade condicional $\mathbb{P}(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)})$ pode ser expressa por:

$$\begin{aligned} \mathbb{P}(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) &= \mathbb{P}\left(\sum_{i=1}^p S_{t,i} + V_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \\ &= \mathbb{P}\left(\sum_{i=2}^p S_{t,i} + V_t + S_{t,1} = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \\ &= \sum_{s_{t,1}=0}^{\min\{y_{t-1}, y_t\}} \mathbb{P}\left(\sum_{i=2}^p S_{t,i} + V_t = y_t - s_{t,1} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \times \mathbb{P}(S_{t,1} = s_{t,1} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}), \quad (4.9) \end{aligned}$$

em que utilizamos a fórmula da convolução (JAMES, 2010, prop. 2.7) para obter a expressão (4.9), referente a probabilidade da soma das v.a.'s independentes $S_{t,1} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$ e $\sum_{i=2}^p S_{t,i} + V_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$. Dessa maneira, utilizando substituições recursivas, obtemos:

$$\begin{aligned} \mathbb{P}(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) &= \sum_{s_{t,1}=0}^{\min\{y_{t-1}, y_t\}} \mathbb{P}\left(\sum_{i=2}^p S_{t,i} + V_t = y_t - s_{t,1} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \\ &\quad \times \mathbb{P}(S_{t,1} = s_{t,1} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \\ &= \sum_{s_{t,1}=0}^{\min\{y_{t-1}, y_t\}} \left[\sum_{s_{t,2}=0}^{\min\{y_{t-2}, y_t - s_{t,1}\}} \mathbb{P}\left(\sum_{i=3}^p S_{t,i} + V_t = y_t - s_{t,1} - s_{t,2} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \right. \\ &\quad \times \mathbb{P}(S_{t,2} = s_{t,2} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \Big] \times \mathbb{P}(S_{t,1} = s_{t,1} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \\ &\quad \vdots \\ &= \sum_{s_{t,1}=0}^{\min\{y_{t-1}, y_t\}} \dots \sum_{s_{t,p}=0}^{\min\{y_{t-(p-1)}, y_t - \sum_{j=1}^{p-1} s_{t,j}\}} \mathbb{P}\left(S_{t,p} + V_t = y_t - \sum_{i=1}^{p-1} s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \\ &\quad \times \prod_{j=1}^{p-1} \mathbb{P}(S_{t,j} = s_{t,j} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \\ &= \sum_{s_{t,1}=0}^{\min\{y_{t-1}, y_t\}} \dots \sum_{s_{t,p}=0}^{\min\{y_{t-p}, y_t - \sum_{j=1}^p s_{t,j}\}} \mathbb{P}\left(V_t = y_t - \sum_{i=1}^p s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \prod_{j=1}^p \mathbb{P}(S_{t,j} = s_{t,j} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \quad (4.10) \end{aligned}$$

Note que a v.a. $S_{t,i}$ depende apenas de Y_{t-i} . Ainda, por definição $S_{t,i}|Y_{t-i} = y_{t-i}$ segue uma distribuição Binomial, com parâmetros y_{t-i} e α_i , se $y_{t-i} > 0$, e $S_{t,i}|Y_{t-i} = y_{t-i}$ segue uma distribuição degenerada em zero, se $y_{t-i} = 0$. Logo:

$$\begin{aligned} \mathbb{P}(S_{t,i} = k | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) &= \mathbb{P}(S_{t,i} = k | Y_{t-i} = y_{t-i}) \\ &= \begin{cases} \binom{y_{t-i}}{k} \alpha_i^k (1 - \alpha_i)^{y_{t-i}-k}, & \text{se } y_{t-i} > 0 \\ 1, & \text{se } y_{t-i} = 0, \end{cases} \end{aligned} \quad (4.11)$$

para $k \in \{0, \dots, y_{t-i}\}$.

Assim, de (4.10), obtemos:

$$\begin{aligned} \mathbb{P}(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) &= \sum_{k_1=0}^{\min\{y_{t-1}, y_t\}} \binom{y_{t-1}}{k_1} \alpha_1^{k_1} (1 - \alpha_1)^{y_{t-1}-k_1} \\ &\times \sum_{k_2=0}^{\min\{y_{t-2}, y_t-k_1\}} \binom{y_{t-2}}{k_2} \alpha_2^{k_2} (1 - \alpha_2)^{y_{t-2}-k_2} \\ &\times \vdots \\ &\times \sum_{k_p=0}^{\min\{y_{t-p}, y_t-\sum_{i=1}^{p-1} k_i\}} \binom{y_{t-p}}{k_p} \alpha_p^{k_p} (1 - \alpha_p)^{y_{t-p}-k_p} \\ &\times \left[\pi \mathbb{I}_{\{y_t-\sum_{i=1}^p k_i=0\}} + (1 - \pi) h_{U_t} \left(y_t - \sum_{i=1}^p k_i | \boldsymbol{\theta} \right) \right], \end{aligned} \quad (4.12)$$

em que $h_{U_t}(\cdot | \boldsymbol{\lambda})$ é a fp da v.a U_t (veja Eq. (2.7) - Cap. 2).

4.4 FUNÇÃO DE VEROSSIMILHANÇA

Seja $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ realizações do processo ZI-INAR(p). Então, a função de verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \pi, \boldsymbol{\lambda})^\top$, dado $\mathbf{y}$ é dada por:

$$\begin{aligned} L(\boldsymbol{\theta} | \mathbf{y}) &= \mathbb{P}(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) \\ &= \mathbb{P}(Y_1 = y_1, \dots, Y_p = y_p) \times \mathbb{P}(Y_{p+1} = y_{p+1} | Y_p = y_p, \dots, Y_1 = y_1) \\ &\quad \times \dots \times \mathbb{P}(Y_n = y_n | Y_{n-1} = y_{n-1}, \dots, Y_{n-p} = y_{n-p}) \\ &= \mathbb{P}(Y_1 = y_1, \dots, Y_p = y_p) \prod_{t=p+1}^n \mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, \dots, Y_{t-p} = y_{t-p}) \end{aligned}$$

Assim como no processo ZI-INAR(1), utilizamos a função de verossimilhança condicional, dada por:

$$\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{y}) &\propto \prod_{t=p+1}^n \mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, \dots, Y_{t-p} = y_{t-p}) \\
&\propto \prod_{t=p+1}^n \left\{ \sum_{k_1=0}^{\min\{y_{t-1}, y_t\}} \binom{y_{t-1}}{k_1} \alpha_1^{k_1} (1 - \alpha_1)^{y_{t-1}-k_1} \right. \\
&\quad \times \sum_{k_2=0}^{\min\{y_{t-2}, y_t-k_1\}} \binom{y_{t-2}}{k_2} \alpha_2^{k_2} (1 - \alpha_2)^{y_{t-2}-k_2} \\
&\quad \times \vdots \\
&\quad \times \sum_{k_p=0}^{\min\{y_{t-p}, y_t-\sum_{i=1}^{p-1} k_i\}} \binom{y_{t-p}}{k_p} \alpha_p^{k_p} (1 - \alpha_p)^{y_{t-p}-k_p} \\
&\quad \left. \times \left[\pi \mathbb{I}_{\{y_t-\sum_{i=1}^p k_i=0\}} + (1 - \pi) h_{U_t} \left(y_t - \sum_{i=1}^p k_i | \boldsymbol{\lambda} \right) \right] \right\}, \quad (4.13)
\end{aligned}$$

A maximização da função (4.13) não é simples de ser obtida. Sendo assim, como no processo ZI-INAR(1), propomos o algoritmo EM como metodologia para estimar os parâmetros.

4.5 ESTIMAÇÃO DOS PARÂMETROS VIA ALGORITMO EM

Nesta seção desenvolvemos o algoritmo EM para estimar os parâmetros do processo ZI-INAR(p). Para isso, definimos as variáveis latentes $\mathbf{S} = (\mathbf{S}_{p+1}, \dots, \mathbf{S}_n)^\top$, sendo $\mathbf{S}_t = (S_{t,1}, \dots, S_{t,p})^\top$, em que $S_{t,i} = \alpha_i \circ Y_{t-i}$, para $i = 1, \dots, p$. Assim, para todo $t = p+1, \dots, n$, a v.a. $S_{t,i} | Y_{t-i} = y_{t-i}$ segue distribuição Binomial com parâmetros y_{t-i} e α_i , se $y_{t-i} > 0$, e $S_{t,i} | Y_{t-i} = y_{t-i}$ segue distribuição degenerada em zero, se $y_{t-i} = 0$. A forma estocástica do processo (4.1) pode ser reescrita como:

$$Y_t = \sum_{i=1}^p S_{t,i} + V_t$$

Assim como no algoritmo EM desenvolvido no processo ZI-INAR(1), consideramos também as variáveis latentes $\mathbf{W} = (W_{p+1}, \dots, W_n)^\top$, para todo $t = p+1, \dots, n$, e $\mathbf{V} = (V_{p+1}, \dots, V_n)^\top$, em que $V_t = Y_t - \sum_{i=1}^p S_{t,i}$, para todo $t = p+1, \dots, n$. Assim, os dados completos são definidos por $\mathbf{Y}_c = (\mathbf{Y}, \mathbf{W}, \mathbf{S}, \mathbf{V})$, em que $\mathbf{Y}$ é o vetor dos dados observados e

$\{\mathbf{W}, \mathbf{S}, \mathbf{V}\}$ é o vetor das variáveis latentes. A função de probabilidade conjunta condicional completa $\mathbb{P}(\mathbf{Y}_{c_t} = \mathbf{y}_{c_t} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)})$ é dada por:

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{c_t} = \mathbf{y}_{c_t} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) &= \mathbb{P}(Y_t = y_t, W_t = w_t, \mathbf{S}_t = \mathbf{s}_t, V_t = v_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \\ &= \mathbb{P}(W_t = w_t, \mathbf{S}_t = \mathbf{s}_t, V_t = v_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \\ &\quad \times \mathbb{P}(Y_t = y_t | W_t = w_t, \mathbf{S}_t = \mathbf{s}_t, V_t = v_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \end{aligned}$$

Dado a restrição $Y_t = \sum_{i=1}^p S_{t,i} + V_t$, temos que o conhecimento das v.a.'s $\mathbf{S}_t$ e V_t determinam o valor que assumirá a v.a. Y_t , o que implica que $\mathbb{P}(Y_t = y_t | W_t = w_t, \mathbf{S}_t = \mathbf{s}_t, V_t = v_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) = 1$. Portanto:

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{c_t} = \mathbf{y}_{c_t} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) &= \mathbb{P}(W_t = w_t, \mathbf{S}_t = \mathbf{s}_t, V_t = v_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \\ &= \mathbb{P}\left(W_t = w_t, \mathbf{S}_t = \mathbf{s}_t, V_t = y_t - \sum_{i=1}^p s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \\ &= \mathbb{P}(\mathbf{S}_t = \mathbf{s}_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \times \mathbb{P}\left(W_t = w_t, V_t = y_t - \sum_{i=1}^p s_{t,i} | \mathbf{S}_t = \mathbf{s}_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \\ &= \mathbb{P}(S_{t,1} = s_{t,1}, \dots, S_{t,p} = s_{t,p} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \times \mathbb{P}\left(W_t = w_t, V_t = y_t - \sum_{i=1}^p s_{t,i}\right) \end{aligned} \quad (4.14)$$

$$\begin{aligned} &= \prod_{i=1}^p \mathbb{P}(S_{t,i} = s_{t,i} | Y_{t-i} = y_{t-i}) \times \mathbb{P}\left(W_t = w_t, V_t = y_t - \sum_{i=1}^p s_{t,i}\right) \\ &= \prod_{i=1}^p \binom{y_{t-i}}{s_{t,i}} \alpha_i^{s_{t,i}} (1 - \alpha_i)^{y_{t-i} - s_{t,i}} \pi^{w_t} \left[(1 - \pi) h_{U_t} \left(y_t - \sum_{i=1}^p s_{t,i} | \boldsymbol{\lambda} \right) \right]^{1-w_t}, \end{aligned} \quad (4.15)$$

em que utilizamos o fato que (W_t, V_t) é independente de $(\mathbf{S}_t, \mathbf{Y}_{t(-p)})$ para obter (4.14). No entanto, a expressão (4.15) é consequência de que $S_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$ é independente de $S_{t,j} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$, para $i \neq j$. Assim, a função de verossimilhança condicional completa é dada por:

$$\begin{aligned} \mathbb{L}_c(\boldsymbol{\theta} | \mathbf{y}_c) &= \mathbb{P}(Y_1 = y_1, \dots, Y_p = y_p) \prod_{t=p+1}^n \mathbb{P}(Y_t = Y_t, \mathbf{S}_t = \mathbf{s}_t, W_t = w_t, V_t = v_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \\ &\propto \prod_{t=p+1}^n \left\{ \prod_{i=1}^p \binom{y_{t-i}}{s_{t,i}} \alpha_i^{s_{t,i}} (1 - \alpha_i)^{y_{t-i} - s_{t,i}} \pi^{w_t} (1 - \pi)^{1-w_t} \times \left[h_{U_t}(y_t - \sum_{i=1}^p s_{t,i} | \boldsymbol{\lambda}) \right]^{1-w_t} \right\}. \end{aligned}$$

Consequentemente, a log-verossimilhança condicional completa é dada por:

$$\begin{aligned}
\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \log(\mathbb{L}_c(\boldsymbol{\theta}|\mathbf{y}_c)) \\
&\propto \sum_{i=1}^p \left\{ \left(\sum_{t=p+1}^n s_{t,i} \right) \log(\alpha_i) \right\} + \sum_{i=1}^p \left\{ \left[\sum_{t=p+1}^n (y_{t-i} - s_{t,i}) \right] \log(1 - \alpha_i) \right\} \\
&+ \sum_{t=p+1}^n w_t \log(\pi) + \sum_{t=p+1}^n (1 - w_t) \log(1 - \pi) \\
&+ \sum_{t=p+1}^n (1 - w_t) \log h_{U_t} \left(y_t - \sum_{i=1}^p s_{t,i} | \boldsymbol{\lambda} \right).
\end{aligned}$$

Portanto, a função Q é dada por:

$$\begin{aligned}
Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) &= E\left(\ell_c(\boldsymbol{\theta}|\mathbf{Y}_c) | \mathbf{Y} = \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\
&\propto \sum_{i=1}^p \left\{ \left(\sum_{t=p+1}^n \hat{s}_{t,i}^{(k)} \right) \log(\alpha_i) \right\} + \sum_{i=1}^p \left\{ \left[\sum_{t=p+1}^n (y_{t-i} - \hat{s}_{t,i}^{(k)}) \right] \log(1 - \alpha_i) \right\} \\
&+ \sum_{t=p+1}^n \hat{w}_t^{(k)} \log(\pi) + \sum_{t=p+1}^n (1 - \hat{w}_t^{(k)}) \log(1 - \pi) + \sum_{t=p+1}^n Q_t^* \left(\boldsymbol{\lambda} | \hat{\boldsymbol{\theta}}^{(k)} \right),
\end{aligned} \tag{4.16}$$

em que $\hat{\boldsymbol{\theta}}^{(k)}$ é a estimativa de $\boldsymbol{\theta}$ na k -ésima iteração do algoritmo e

- $\hat{s}_{t,i}^{(k)} = E\left(S_{t,i} | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right);$
- $\hat{w}_t^{(k)} = E\left(W_t | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right);$
- $Q_t^* \left(\boldsymbol{\lambda} | \hat{\boldsymbol{\theta}}^{(k)} \right) = E\left((1 - W_t) \log h\left(y_t - \sum_{i=1}^p S_{t,i}\right) | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right).$

Assim, em cada iteração, no passo E do algoritmo EM é obtido $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$, o qual será maximizado no passo M. A seguir, mostramos os dois passos do algoritmo.

Passo E:

Note que $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ é obtido pelo conhecimento de $\hat{s}_{t,i}^{(k)}$, $\hat{w}_t^{(k)}$ e $Q_t^* \left(\boldsymbol{\lambda} | \hat{\boldsymbol{\theta}}^{(k)} \right)$.

Para a k -ésima iteração do algoritmo EM, obtemos $\hat{s}_{t,i}^{(k)}$ por meio do resultado:

$$\begin{aligned}
\mathbb{P}\left(S_{t,i} = s_{t,i} | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \frac{\mathbb{P}\left(S_{t,i} = s_{t,i}, Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(S_{t,i} = s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(Y_t = y_t | S_{t,i} = s_{t,i}, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\binom{y_{t-i}}{s_{t,i}} (\hat{\alpha}_i^{(k)})^{s_{t,i}} (1 - \hat{\alpha}_i^{(k)})^{y_{t-i}-s_{t,i}} \times \mathbb{P}\left(\sum_{i=1}^p S_{t,i} + V_t = y_t | S_{t,i} = s_{t,i}, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}, \tag{4.17}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\binom{y_{t-i}}{s_{t,i}} (\hat{\alpha}_i^{(k)})^{s_{t,i}} (1 - \hat{\alpha}_i^{(k)})^{y_{t-i}-s_{t,i}} \times \mathbb{P}\left(\sum_{j \neq i}^p S_{t,j} + V_t = y_t - s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}, \tag{4.18}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\binom{y_{t-i}}{s_{t,i}} (\hat{\alpha}_i^{(k)})^{s_{t,i}} (1 - \hat{\alpha}_i^{(k)})^{y_{t-i}-s_{t,i}} \times \mathbb{P}\left(Y_{t,i} = y_t - s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}, \tag{4.19}
\end{aligned}$$

em que $Y_{t,i} = \sum_{j \neq i}^p S_{t,j} + V_t$. A expressão (4.17) é consequência do resultado obtido em (4.11). No entanto, a expressão (4.18) é obtida considerando que, condicionado a $\mathbf{y}_{t(-p)}$, as v.a.'s $\sum_{j \neq i}^p S_{t,j} + V_t$ e $S_{t,i}$ são independentes.

Agora, note que $\mathbb{P}\left(Y_{t,i} = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)$ representa a probabilidade condicional de um processo ZI-INAR(p), com vetor de parâmetros $\hat{\boldsymbol{\theta}}^{*(k)} = (\hat{\boldsymbol{\alpha}}^{*(k)}, \hat{\pi}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)})$, em que $\hat{\boldsymbol{\alpha}}^{*(k)} = (\hat{\alpha}_1^{(k)}, \dots, \hat{\alpha}_{i-1}^{(k)}, 0, \hat{\alpha}_{i+1}^{(k)}, \dots, \hat{\alpha}_p^{(k)})$. De fato, reescrevendo $Y_{t,i}$, obtemos:

$$Y_{t,i} = \sum_{j \neq i}^p S_{t,j} + V_t = \sum_{j \neq i}^p \alpha_j \circ Y_{t-j} + 0 \circ Y_{t-i} + V_t, \tag{4.20}$$

em que $0 \circ Y_{t-i} = 0$ (veja Proposição 2.1 – Cap. 2). Assim, a expressão (4.19) pode ser calculada utilizando a fórmula (4.12) e $\hat{s}_{t,i}^{(k)}$ é obtido por:

$$\begin{aligned}
\hat{s}_{t,i}^{(k)} &= E\left(S_{t,i} | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \frac{\sum_{s_{t,i}=0}^{\min\{y_{t-i}, y_t\}} s_{t,i} \binom{y_{t-i}}{s_{t,i}} (\hat{\alpha}_i^{(k)})^{s_{t,i}} (1 - \hat{\alpha}_i^{(k)})^{y_{t-i}-s_{t,i}} \times \mathbb{P}\left(Y_{t,i} = y_t - s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}.
\end{aligned}$$

E em cada passo, $\hat{w}_t^{(k)}$ é obtido como:

$$\begin{aligned}
\hat{w}_t^{(k)} &= E\left(W_t | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \frac{\mathbb{P}\left(W_t = 1, Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(W_t = 1 | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(Y_t = y_t | W_t = 1, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(W_t = 1 | \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(\sum_{i=1}^p S_{t,i} + V_t = y_t | W_t = 1, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \quad (4.21)
\end{aligned}$$

$$= \frac{\hat{\pi}^{(k)} \times \mathbb{P}\left(\sum_{i=1}^p S_{t,i} + V_t^* = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \quad (4.22)$$

$$= \frac{\hat{\pi}^{(k)} \times \mathbb{P}\left(\sum_{i=1}^p S_{t,i} = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}. \quad (4.23)$$

em que V_t^* é uma v.a. degenerada em zero. A expressão (4.21) é obtido pela independência entre W_t e $\mathbf{Y}_{t(-p)}$, e (4.22) é consequência de que as v.a.'s $V_t | W_t = 1$ e V_t^* são iguais em distribuição. Agora, note que se $\pi = 1$, então $\mathbb{P}(V_t = 0) = 1$. Portanto, substituindo π por “1” na fórmula (4.12), temos que:

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^p S_{t,i} = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \mathbb{P}\left(\sum_{i=1}^p S_{t,i} + V_t^* = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) \\
&= \sum_{k_1=0}^{\min\{y_{t-1}, y_t\}} \binom{y_{t-1}}{k_1} \alpha_1^{k_1} (1 - \alpha_1)^{y_{t-1} - k_1} \\
&\times \sum_{k_2=0}^{\min\{y_{t-2}, y_t - k_1\}} \binom{y_{t-2}}{k_2} \alpha_2^{k_2} (1 - \alpha_2)^{y_{t-2} - k_2} \\
&\times \vdots \\
&\times \sum_{k_{p-1}=0}^{\min\{y_{t-(p-1)}, y_t - \sum_{i=1}^{p-2} k_i\}} \binom{y_{t-(p-1)}}{k_{p-1}} \alpha_{p-1}^{k_{p-1}} (1 - \alpha_{p-1})^{y_{t-(p-1)} - k_{p-1}} \\
&\times \binom{y_{t-p}}{y_t - \sum_{i=1}^{p-1} k_i} \alpha_p^{y_t - \sum_{i=1}^{p-1} k_i} (1 - \alpha_p)^{y_{t-p} - (y_t - \sum_{i=1}^{p-1} k_i)} \quad (4.24)
\end{aligned}$$

Assim, obtemos $\hat{w}_t$ pela fórmula (4.23) utilizando os resultados (4.12) e (4.24).

A função $Q_t^* \left(\boldsymbol{\lambda} | \hat{\boldsymbol{\theta}}^{(k)} \right)$ é caracterizada pela distribuição Inflacionada de Zeros considerada para as inovações. Sendo assim, apresentamos o algoritmo EM para os casos propostos.

Passo M:

Na Seção 4.1, vimos que o processo ZI-INAR(p) será estacionário se o vetor $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)$ pertencer à região do $\mathbb{R}^p$ tal que $\alpha_i \in [0, 1)$ e $\sum_{i=1}^p \alpha_i < 1$, para $i = 1, \dots, p$. Denotaremos tal região de região de estacionariedade de $\boldsymbol{\alpha}$. Assim, para garantir a estacionariedade do processo, a estimativa de $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \pi, \boldsymbol{\lambda})^\top$ é obtida pela maximização da função $Q \left(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)} \right)$, a qual deve estar restrita à região de estacionariedade de $\boldsymbol{\alpha}$. Com o intuito de contornar este problema, propomos o seguinte teorema:

Teorema 4.1. *Seja $\Delta : (\alpha_1, \dots, \alpha_p) \rightarrow (\beta_1, \dots, \beta_p)$ uma transformação em $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)$ tal que:*

$$\Delta_i(\alpha_1, \dots, \alpha_p) = \beta_i = \frac{\alpha_i}{1 - \sum_{j \neq i} \alpha_j}, \quad j = 1, \dots, p. \quad (4.25)$$

Se $\alpha_i \in [0, 1)$ e $\sum_{i=1}^p \alpha_i < 1$, então temos que:

- (i) o conjunto imagem da transformação Δ é um cubo p -dimensional limitado no intervalo $[0, 1)^p$.
- (ii) a transformação Δ admite inversa dada por:

$$\alpha_i = \frac{\frac{\beta_i}{1 - \beta_i} \prod_{l=1}^p (1 - \beta_l)}{1 - \sum_{j=1}^{p-1} \sum_{k=j+1}^p \beta_j \frac{\beta_k}{1 - \beta_k} \prod_{l=j+1}^p (1 - \beta_l)}, \quad i = 1, \dots, p. \quad (4.26)$$

Prova. Ver Apêndice C.

A partir do resultado (4.26) podemos reescrever a equação (4.16) da seguinte forma:

$$\begin{aligned}
Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) &\propto \sum_{i=1}^p \left(\sum_{t=p+1}^n \hat{s}_{t,i}^{(k)} \right) \log\left(\Delta_i^{-1}(\boldsymbol{\beta})\right) \\
&+ \sum_{i=1}^p \left[\sum_{t=p+1}^n \left(y_{t-i} - \hat{s}_{t,i}^{(k)} \right) \right] \log\left(1 - \Delta_i^{-1}(\boldsymbol{\beta})\right) \\
&+ \sum_{t=p+1}^n \hat{w}_t^{(k)} \log(\pi) + \sum_{t=p+1}^n \left(1 - \hat{w}_t^{(k)}\right) \log(1 - \pi) + \sum_{t=p+1}^n Q_t^*\left(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}}^{(k)}\right),
\end{aligned} \tag{4.27}$$

em que $\Delta_i^{-1}(\boldsymbol{\beta}) = \alpha_i$. As primeiras derivadas parciais da função (4.27) são dadas por:

$$\begin{aligned}
\frac{\partial Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)}{\partial \beta_i} &= \sum_{i=1}^p \left(\sum_{t=p+1}^n \hat{s}_{t,i}^{(k)} \right) \frac{1}{\Delta_i^{-1}(\boldsymbol{\beta})} \times \frac{\partial \Delta_i^{-1}(\boldsymbol{\beta})}{\partial \beta_i} \\
&- \sum_{i=1}^p \left[\sum_{t=p+1}^n \left(y_{t-i} - \hat{s}_{t,i}^{(k)} \right) \right] \frac{1}{1 - \Delta_i^{-1}(\boldsymbol{\beta})} \times \frac{\partial \Delta_i^{-1}(\boldsymbol{\beta})}{\partial \beta_i}, \quad i = 1, \dots, p
\end{aligned} \tag{4.28}$$

$$\frac{\partial Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)}{\partial \pi} = \frac{\sum_{t=p+1}^n \hat{w}_t^{(k)}}{\pi} - \frac{\sum_{t=p+1}^n (1 - \hat{w}_t^{(k)})}{1 - \pi} \tag{4.29}$$

$$\frac{\partial Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)}{\partial \boldsymbol{\lambda}} = \sum_{t=p+1}^n \frac{\partial Q_t^*(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})}{\partial \boldsymbol{\lambda}} \tag{4.30}$$

Note que as derivadas parciais de $Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ com relação a β_i , π e $\boldsymbol{\lambda}$ dependem apenas de $\boldsymbol{\beta}$, π e $\boldsymbol{\lambda}$, respectivamente. Assim, podemos obter os estimadores de $\hat{\boldsymbol{\alpha}}^{(k)}$, $\hat{\pi}^{(k)}$ e $\hat{\boldsymbol{\lambda}}^{(k)}$ separadamente, como podem ser vistas a seguir.

Estimadores de $\boldsymbol{\beta}$ e $\boldsymbol{\alpha}$

Neste caso, não é possível obter de forma analítica, a maximização de $Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ com relação aos β'_i s. Assim, a atualização de $\hat{\boldsymbol{\beta}}^{(k)}$ é dada por:

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \arg \max_{\boldsymbol{\beta}} \left\{ Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) \right\}, \tag{4.31}$$

e $\hat{\alpha}_i^{(k+1)}$ é obtido por:

$$\begin{aligned}
\hat{\alpha}_i^{(k+1)} &= \frac{\prod_{l=1}^p (1 - \hat{\beta}_l^{(k+1)})}{1 - \hat{\beta}_i^{(k+1)}}, \quad i = 1, \dots, p. \\
&\frac{\hat{\beta}_j^{(k+1)} \hat{\beta}_k^{(k+1)} \prod_{l=j+1}^p (1 - \hat{\beta}_l^{(k+1)})}{1 - \sum_{j=1}^{p-1} \sum_{k=j+1}^p \frac{\hat{\beta}_j^{(k+1)} \hat{\beta}_k^{(k+1)} \prod_{l=j+1}^p (1 - \hat{\beta}_l^{(k+1)})}{1 - \hat{\beta}_k^{(k+1)}}}
\end{aligned} \tag{4.32}$$

Estimadores de π e λ

Utilizando os resultados obtidos em (4.29) e (4.30), temos que:

$$\hat{\pi}^{(k+1)} = \frac{\sum_{t=p+1}^n \hat{w}_t}{n-p} \quad (4.33)$$

$$\hat{\lambda}^{(k+1)} = \arg \max_{\lambda} \left\{ \sum_{t=p+1}^n Q_t^*(\lambda | \hat{\theta}^{(k)}) \right\}. \quad (4.34)$$

Assim, o algoritmo EM pode ser resumido da seguinte maneira:

- **Passo E:** Dado $\hat{\theta}^{(k)}$, para $t = p+1, \dots, n$ e $i = 1, \dots, p$, calculamos $\hat{s}_{t,i}^{(k)}$, $\hat{w}_t^{(k)}$ e $Q_t^*(\lambda | \theta^{(k)})$.
- **Passo M:** Atualize $\hat{\theta}^{(k)}$ pela maximização $Q(\theta | \hat{\theta}^{(k)})$ sobre θ , o que implica nas expressões (4.31)-(4.34) para $\hat{\beta}^{(k+1)}$, $\hat{\alpha}^{(k+1)}$, $\hat{\pi}^{(k+1)}$ e $\hat{\lambda}^{(k+1)}$, respectivamente.

Os passos E e M são iterados até que algum critério de convergência seja satisfeito. Assim como no algoritmo EM para o processo ZI-INAR(1), utilizamos o critério de parada baseado na aceleração de Aitken (MCLACHLAN; KRISHNAN, 2008).

Em seguida, mostramos como atualizar $\hat{\lambda}^{(k)}$ em cada passo do algoritmo, para os casos particulares ZIP-INAR(p), ZINB-INAR(p) e ZIPIG-INAR(p). Para os três casos, será necessário no passo E, o cálculo do valor esperado $\widehat{b_{ts_{t,i}}} = E\left(B_t S_{t,i} | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)}\right)$, sendo $B_t = 1 - W_t$. Por analogia ao cálculo de $\widehat{b_{ts_t}}$ obtido no algoritmo EM do processo ZI-INAR(1), temos que:

$$\begin{aligned} \widehat{b_{ts_{t,i}}} &= E\left(B_t S_{t,i} | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)}\right) \\ &= \mathbb{P}\left(W_t = 0 | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)}\right) \times E\left(S_{t,i} | W_t = 0, Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)}\right), \end{aligned} \quad (4.35)$$

em que

$$\mathbb{P}\left(W_t = 0 | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)}\right) = \frac{\left(1 - \hat{\pi}^{(k)}\right) \mathbb{P}\left(Y_t = y_t | W_t = 0, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)}\right)}. \quad (4.36)$$

Para calcular o valor esperado na expressão (4.35), temos que:

$$\begin{aligned}
\mathbb{P}\left(S_{t,i} = s_{t,i} | W_t = 0, Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \frac{\mathbb{P}\left(S_{t,i} = s_{t,i}, Y_t = y_t | W_t = 0, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | W_t = 0, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(S_{t,i} = s_{t,i} | W_t = 0, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(Y_t = y_t | S_{t,i} = s_{t,i}, W_t = 0, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | W_t = 0, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \\
&= \frac{\mathbb{P}\left(S_{t,i} = s_{t,i} | \mathbf{Y}_{t,i} = \mathbf{y}_{t,i}, \hat{\boldsymbol{\theta}}^{(k)}\right) \times \mathbb{P}\left(\sum_{j=1}^p S_{t,j} + U_t = y_t | S_{t,i} = s_{t,i}, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | W_t = 0, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \quad (4.37)
\end{aligned}$$

$$= \frac{\binom{y_{t-i}}{s_{t,i}} (\hat{\alpha}^{(k)})^{s_{t,i}} (1 - \hat{\alpha}^{(k)})^{y_{t-i} - s_{t,i}} \times \mathbb{P}\left(\sum_{j \neq i}^p S_{t,j} + U_t = y_t - s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | W_t = 0, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}, \quad (4.38)$$

em que (4.37) é consequência de que $S_{t,i}$ e W_t são v.a's independentes e de que, dado $W_t = 0$, temos que $V_t = U_t$. Por outro lado, a expressão (4.38) é obtida pelo resultado (4.11) e pela independência das v.a's $\sum_{j \neq i}^p S_{t,j} + U_t$ e $S_{t,i}$. Definindo $Y_{t,i}^* = \sum_{j \neq i}^p S_{t,j} + U_t$, obtemos:

$$\begin{aligned}
E\left(S_{t,i} | W_t = 0, Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right) &= \\
&= \frac{\sum_{s_{t,i}=0}^{\min\{y_{t-i}, y_t\}} s_{t,i} \binom{y_{t-i}}{s_{t,i}} (\hat{\alpha}_i^{(k)})^{s_{t,i}} (1 - \hat{\alpha}_i^{(k)})^{y_{t-i} - s_{t,i}} \times \mathbb{P}\left(Y_{t,i}^* = y_t - s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | W_t = 0, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)} \quad (4.39)
\end{aligned}$$

Note que, podemos simplificar a expressão (4.35) utilizando os resultados (4.36) e (4.39), o que implica que $\widehat{b_t s_{t,i}}$ pode ser expresso como:

$$\widehat{b_t s_{t,i}} = \frac{(1 - \hat{\pi}^{(k)}) \sum_{s_{t,i}=0}^{\min\{y_{t-i}, y_t\}} s_{t,i} \binom{y_{t-i}}{s_{t,i}} (\hat{\alpha}_i^{(k)})^{s_{t,i}} (1 - \hat{\alpha}_i^{(k)})^{y_{t-i} - s_{t,i}} \times \mathbb{P}\left(Y_{t,i}^* = y_t - s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}{\mathbb{P}\left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)}, \quad (4.40)$$

Agora, note que por analogia a (4.20), $\mathbb{P}\left(Y_{t,i}^* = y_t - s_{t,i} | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)}\right)$ representa a probabilidade condicional de um processo ZI-INAR(p), com inovações $\{U_t\}_{t \in \mathbb{Z}}$ e vetor de parâmetros $\hat{\boldsymbol{\theta}}^{*(k)} = (\hat{\boldsymbol{\alpha}}^{*(k)}, \hat{\boldsymbol{\lambda}}^{(k)})$, em que $\hat{\boldsymbol{\alpha}}^{*(k)} = (\hat{\alpha}_1^{(k)}, \dots, \hat{\alpha}_{i-1}^{(k)}, 0, \hat{\alpha}_{i+1}^{(k)}, \dots, \hat{\alpha}_p^{(k)})$. Assim, podemos obter (4.40) utilizando o resultado (4.12).

1. Caso ZIP:

Se $V_t \sim \text{ZIP}(\pi, \lambda)$, então $h_{U_t}(\cdot | \lambda)$ é a fp de uma v.a Poisson(λ). Assim, temos que:

$$(1 - w_t) \log \left\{ h_{U_t} \left(y_t - \sum_{i=1}^p s_{t,i} | \lambda \right) \right\} \propto -(1 - w_t) \lambda + (1 - w_t) \left(y_t - \sum_{i=1}^p s_{t,i} \right) \log(\lambda).$$

Logo,

$$\begin{aligned} \sum_{t=p+1}^n Q_t^*(\mu | \theta^{(k)}) &\propto \sum_{t=p+1}^n E \left((1 - W_t) \log h \left(y_t - \sum_{i=1}^p S_{t,i} \right) | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)} \right) \\ &\propto -\lambda \sum_{t=p+1}^n (1 - \hat{w}_t^{(k)}) + \log(\lambda) \sum_{t=p+1}^n (1 - \hat{w}_t^{(k)}) y_t + \log(\lambda) \sum_{t=p+1}^n \sum_{i=1}^p \widehat{b_{ts_{t,i}}}^{(k)} \end{aligned} \quad (4.41)$$

Maximizando a expressão (4.41), com relação a λ , temos que:

$$\hat{\lambda}^{(k+1)} = \frac{\sum_{t=p+1}^n (1 - \hat{w}_t^{(k)}) y_t - \sum_{t=p+1}^n \sum_{i=1}^p \widehat{b_{ts_{t,i}}}^{(k)}}{\sum_{t=p+1}^n (1 - \hat{w}_t^{(k)})}.$$

Note que a atualização de $\hat{\lambda}^{(k)}$ pertence ao espaço paramétrico de λ . De fato, temos que:

$$\begin{aligned} \sum_{t=p+1}^n (1 - \hat{w}_t^{(k)}) y_t - \sum_{t=p+1}^n \sum_{i=1}^p \widehat{b_{ts_{t,i}}}^{(k)} &= \sum_{t=2}^n \left(1 - E \left(W_t | \mathbf{y}, \hat{\theta}^{(k)} \right) \right) y_t - \sum_{t=p+1}^n \sum_{i=1}^p E \left(B_t S_{t,i} | \mathbf{y}, \hat{\theta}^{(k)} \right) \\ &= \sum_{t=p+1}^n E \left(B_t Y_t | \mathbf{y}, \hat{\theta}^{(k)} \right) - \sum_{t=p+1}^n \sum_{i=1}^p E \left(B_t S_{t,i} | \mathbf{y}, \hat{\theta}^{(k)} \right) \\ &= \sum_{t=p+1}^n E \left(B_t \left(Y_t - \sum_{i=1}^p S_{t,i} \right) | \mathbf{y}, \hat{\theta}^{(k)} \right) \\ &= \sum_{t=p+1}^n E \left(B_t V_t | \mathbf{y}, \hat{\theta}^{(k)} \right) > 0. \end{aligned}$$

2. Caso ZINB:

Se $V_t \sim \text{ZINB}(\pi, \mu, \phi)$, então $\lambda = (\mu, \phi)$ e $h_{U_t}(\cdot | \lambda)$ é a fp de uma v.a NB(μ, ϕ). Assim, temos que:

$$\begin{aligned} (1 - w_t) \log \left\{ h_{U_t} \left(y_t - \sum_{i=1}^p s_{t,i} | \lambda \right) \right\} &\propto (1 - w_t) \left\{ \log \Gamma \left(y_t - \sum_{i=1}^p s_{t,i} + \phi \right) - \log(\Gamma(\phi)) \right. \\ &\quad \left. + \left(y_t - \sum_{i=1}^p s_{t,i} \right) [\log(\mu) - \log(\mu + \phi)] + \phi [\log(\phi) - \log(\mu + \phi)] \right\}. \end{aligned}$$

Logo,

$$\begin{aligned}
\sum_{t=p+1}^n Q_t^*(\mu, \phi | \boldsymbol{\theta}^{(k)}) &\propto \sum_{t=p+1}^n E \left(B_t \log h \left(y_t - \sum_{i=1}^p S_{t,i} \right) | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)} \right) \\
&\propto \sum_{t=p+1}^n \widehat{g_t(\phi)}^{(k)} + [\log(\mu) - \log(\mu + \phi)] \\
&\times \left(\sum_{t=p+1}^n (1 - \hat{w}_t^{(k)}) y_t - \sum_{t=p+1}^n \sum_{i=1}^p \widehat{b_t s_{t,i}}^{(k)} \right) \\
&+ [-\log(\Gamma(\phi)) + \phi [\log(\phi) - \log(\mu + \phi)]] \sum_{t=p+1}^n (1 - \hat{w}_t^{(k)}),
\end{aligned}$$

em que $\widehat{g_t(\phi)}^{(k)} = E \left(B_t \log \Gamma \left(y_t - \sum_{i=1}^p S_{t,i} + \phi \right) | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\boldsymbol{\theta}}^{(k)} \right)$.

Para obter as atualizações de $\hat{\mu}^{(k)}$ e $\hat{\phi}^{(k)}$, utilizamos o algoritmo ECME, proposto por Liu e Rubin (1994). O algoritmo ECME é uma extensão do algoritmo ECM (MENG; RUBIN, 1993), o qual substitui as etapas de maximização condicional da log-verossimilhança completa no passo CM, por etapas de maximização condicional da log-verossimilhança incompleta.

Assim, o passo CME para o processo ZINB-INAR(p) é dado pelas seguintes etapas:

1. Atualize $\hat{\mu}^{(k)}$ pela maximização de $\sum_{t=p+1}^n \left\{ Q_t^*(\mu, \hat{\phi}^{(k)} | \hat{\boldsymbol{\theta}}^{(k)}) \right\}$ com relação a μ .
2. Atualize $\hat{\phi}^{(k)}$ pela maximização de $\ell \left(\hat{\boldsymbol{\alpha}}^{(k)}, \hat{\pi}^{(k)}, \hat{\mu}^{(k+1)}, \phi | \mathbf{y} \right)$ com relação a ϕ , em que $\ell(\boldsymbol{\theta} | \mathbf{y}) = \log L(\boldsymbol{\theta} | \mathbf{y})$, e $L(\boldsymbol{\theta} | \mathbf{y})$ é definido em (4.13).

Note que no passo 2 do algoritmo substituímos a maximização da função de log-verossimilhança completa pela maximização da função de log-verossimilhança marginal e o cálculo de $\widehat{g_t(\phi)}^{(k)}$ não é mais necessário. Conforme mencionado por Liu e Rubin (1994), em geral, o algoritmo ECME é mais rápido do que o algoritmo EM e ECM em termos de número de iterações, podendo em algumas situações, ser mais rápido também com relação ao tempo.

Assim, a atualização de $\hat{\mu}^{(k)}$ e $\hat{\phi}^{(k)}$ são dadas por:

$$\hat{\mu}^{(k+1)} = \frac{\sum_{t=p+1}^n (1 - \hat{w}_t^{(k)}) y_t - \sum_{t=p+1}^n \sum_{i=1}^p \widehat{b_t s_{t,i}}^{(k)}}{\sum_{t=p+1}^n (1 - \hat{w}_t^{(k)})}.$$

$$\hat{\phi}^{(k+1)} = \arg \max_{\phi} \left\{ \ell \left(\hat{\boldsymbol{\alpha}}^{(k)}, \hat{\pi}^{(k)}, \hat{\mu}^{(k+1)}, \phi | \mathbf{y} \right) \right\}$$

A solução de $\hat{\phi}^{(k+1)}$ é obtida utilizando a função “optim” do software R ((R Core Team, 2019), 2019).

3. Caso ZIPIG:

Se $V_t \sim \text{ZIPIG}(\pi, \mu, \phi)$, então $\lambda = (\mu, \phi)$. Para desenvolver o algoritmo EM, consideramos a representação estocástica da v.a. V_t em termos da variável latente Z_t , definida no algoritmo EM do processo ZIPIG-INAR(1). Sendo assim, os dados completos será $\mathbf{Y}_c = (\mathbf{Y}, \mathbf{W}, \mathbf{S}, \mathbf{Z})$. Assumindo que $y_t = \sum_{i=1}^n s_t + v_t$, e utilizando o mesmo argumento de (3.19), temos que a fdp conjunta de $\mathbf{Y}_c$ dado $\mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}$ é

$$f(\mathbf{y}_{c_t} | \mathbf{y}_{t(-p)}) = f(y_t, \mathbf{s}_t, w_t, v_t, z_t | \mathbf{y}_{t(-p)}) = f(y_t, s_t, w_t, z_t | \mathbf{y}_{t(-p)}).$$

Portanto,

$$\begin{aligned} f(\mathbf{y}_{c_t} | \mathbf{y}_{t(-p)}) &= f(y_t, \mathbf{s}_t, w_t, z_t | \mathbf{y}_{t(-p)}) \\ &= f(y_t, w_t, z_t | \mathbf{s}_t, \mathbf{y}_{t(-p)}) \times \mathbb{P}(\mathbf{S}_t = \mathbf{s}_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \\ &= f(w_t, z_t) \times \mathbb{P}(Y_t = y_t | \mathbf{S}_t = \mathbf{s}_t, W_t = w_t, Z_t = z_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \\ &\quad \times \mathbb{P}(\mathbf{S}_t = \mathbf{s}_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) \end{aligned} \quad (4.42)$$

$$\begin{aligned} &= f(z_t | w_t) \times \mathbb{P}(W_t = w_t) \times \mathbb{P}\left(V_t = y_t - \sum_{i=1}^p s_{t,i} | W_t = w_t, Z_t = z_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \\ &\quad \times \prod_{i=1}^p \mathbb{P}(S_{t,i} = s_{t,i} | Y_{t-i} = y_{t-i}) \end{aligned} \quad (4.43)$$

$$\begin{aligned} &= \left\{ \frac{\phi^{1/2}}{\sqrt{2\pi}} z_t^{-3/2} \exp\left\{-\frac{\phi(z_t - 1)^2}{2z_t}\right\} \right\}^{1-w_t} \pi^{w_t} (1 - \pi)^{1-w_t} \left[\frac{e^{-\mu z_t} (\mu z_t)^{y_t - \sum_{i=1}^p s_{t,i}}}{\left(y_t - \sum_{i=1}^p s_{t,i}\right)!} \right]^{1-w_t} \\ &\quad \times \prod_{i=1}^p \binom{y_{t-i}}{s_t} \alpha_i^{s_t} (1 - \alpha_i)^{y_{t-i} - s_{t,i}}, \end{aligned} \quad (4.44)$$

em que (4.42) é obtido pela independência de (W_t, Z_t) e $(\mathbf{S}_t, \mathbf{Y}_{t(-p)})$, e a expressão (4.43) é obtida pelo seguinte resultado:

$$\begin{aligned}
\mathbb{P}(Y_t = y_t | \mathbf{S}_t = \mathbf{s}_t, W_t = w_t, Z_t = z_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}) &= \\
&= \mathbb{P}\left(\sum_{i=1}^p S_{t,i} + V_t = y_t | \mathbf{S}_t = \mathbf{s}_t, W_t = w_t, Z_t = z_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \\
&= \mathbb{P}\left(V_t = y_t - \sum_{i=1}^p s_{t,i} | W_t = w_t, Z_t = z_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}\right) \quad (4.45)
\end{aligned}$$

$$= \mathbb{P}\left(V_t = y_t - \sum_{i=1}^p s_{t,i} | W_t = w_t, Z_t = z_t\right). \quad (4.46)$$

O resultado (4.45) é consequência da suposição de independência de $\sum_{i=1}^p S_{t,i}$ e V_t , e (4.46) é consequência da independência de V_t e $\mathbf{Y}_{t(-p)}$.

Assim, de (4.44), a função de verossimilhança completa é dada por:

$$\begin{aligned}
\mathbb{L}_c(\boldsymbol{\theta} | \mathbf{y}_c) &\propto \prod_{t=p+1}^n \left\{ \pi^{w_t} (1-\pi)^{1-w_t} \prod_{i=1}^p \binom{y_{t-i}}{s_{t,i}} \alpha_i^{s_{t,i}} (1-\alpha_i)^{y_{t-i}-s_{t,i}} \right. \\
&\quad \times \left. \left[\frac{e^{-\mu z_t} (\mu z_t)^{y_t - \sum_{i=1}^p s_{t,i}}}{\left(y_t - \sum_{i=1}^p s_{t,i}\right)!} \right]^{1-w_t} \left\{ \phi^{1/2} z_t^{-3/2} \exp\left\{-\frac{\phi(z_t-1)^2}{2z_t}\right\} \right\}^{1-w_t} \right\}.
\end{aligned}$$

Então, a função de log-verossimilhança condicional completa é dada por

$$\begin{aligned}
\ell_c(\boldsymbol{\theta} | \mathbf{y}_c) &= \log(\mathbb{L}_c(\boldsymbol{\theta} | \mathbf{y}_c)) \\
&\propto \sum_{t=p+1}^n w_t \log(\pi) + \sum_{t=2}^n (1-w_t) \log(1-\pi) \\
&\quad + \sum_{i=1}^p \left(\sum_{t=p+1}^n s_{t,i} \right) \log(\alpha_i) + \sum_{i=1}^p \left[\sum_{t=p+1}^n (y_{t-i} - s_{t,i}) \right] \log(1-\alpha_i) \\
&\quad + \sum_{t=p+1}^n (1-w_t) \left[-\mu z_t + \left(y_t - \sum_{i=1}^p s_{t,i} \right) \log(\mu) - \frac{\phi}{2} z_t - \frac{\phi}{2} z_t^{-1} + \phi + \frac{1}{2} \log(\phi) \right].
\end{aligned}$$

Assim, a função $Q\left(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}\right)$ pode ser expressa na forma reduzida como

$$\begin{aligned}
Q\left(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}\right) &\propto \sum_{t=p+1}^n \hat{w}_t^{(k)} \log(\pi) + \sum_{t=p+1}^n \left(1 - \hat{w}_t^{(k)}\right) \log(1-\pi) \\
&\quad + \sum_{i=1}^p \left(\sum_{t=p+1}^n \hat{s}_{t,i}^{(k)} \right) \log(\alpha_i) + \sum_{i=1}^p \left[\sum_{t=p+1}^n \left(y_{t-i} - \hat{s}_{t,i}^{(k)} \right) \right] \log(1-\alpha_i) \\
&\quad + \sum_{t=p+1}^n Q_t^{**}(\boldsymbol{\lambda} | \boldsymbol{\theta}^{(k)}),
\end{aligned}$$

em que

$$Q_t^{**}(\lambda|\theta^{(k)}) \propto -\mu \sum_{t=p+1}^n \widehat{b_t z_t}^{(k)} - \log(\mu) \sum_{i=p}^n \sum_{t=p+1}^n \widehat{b_t s_{t,i}}^{(k)} + \log(\mu) \sum_{t=p+1}^n (1 - \widehat{w_t}^{(k)}) y_t \\ + \frac{\log(\phi)}{2} \sum_{t=p+1}^n (1 - \widehat{w_t}^{(k)}) + \phi \sum_{t=p+1}^n (1 - \widehat{w_t}^{(k)}) - \frac{\phi}{2} \sum_{t=p+1}^n \widehat{b_t z_t}^{(k)} - \frac{\phi}{2} \sum_{t=p+1}^n \widehat{b_t / z_t}^{(k)}$$

Assim, no passo E do algoritmo, é obtido $\hat{s}_{t,i}, \hat{w}_t$ e a função $Q_t^{**}(\mu, \phi|\hat{\theta}^{(k)})$, em que esta última é completamente determinada pelo conhecimento de $\widehat{b_t s_{t,i}}^{(k)}, \widehat{b_t z_t}^{(k)}$ e $\widehat{b_t / z_t}^{(k)}$, sendo $\widehat{b_t z_t}^{(k)} = E(B_t Z_t | \mathbf{y}, \theta^{(k)})$ e $\widehat{b_t / z_t}^{(k)} = E(B_t Z_t^{-1} | \mathbf{y}, \hat{\theta}^{(k)})$. As esperanças $\widehat{b_t z_t}^{(k)}$ e $\widehat{b_t / z_t}^{(k)}$ podem ser expressas por (veja demonstração no Apêndice B):

$$\widehat{b_t z_t}^{(k)} = E \left[E \left(B_t Z_t | \mathbf{S}_t, Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\lambda}^{(k)} \right) \right] \\ = \frac{\min \left\{ \sum_{i=1}^p y_{t-i}, y_t \right\} \left(1 - \hat{\pi}^{(k)} \right) (y_t - s + 1) h \left(y_t - s + 1 | \hat{\lambda}^{(k)} \right) \mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)} \right)}{\hat{\mu}^{(k)} \left[\hat{\pi}^{(k)} I_{\{y_t - s = 0\}} + (1 - \hat{\pi}^{(k)}) h_{U_t} \left(y_t - s | \hat{\lambda}^{(k)} \right) \right]}$$

e

$$\widehat{b_t / z_t}^{(k)} = E \left[E \left(B_t Z_t^{-1} | \mathbf{S}_t, Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\lambda}^{(k)} \right) \right] \\ = \sum_{s=0}^{\min \left\{ \sum_{i=1}^p y_{t-i}, y_t \right\}} \left\{ \frac{\hat{\mu}^{(k)} h \left(y_t - s - 1 | \hat{\lambda}^{(k)} \right) \times \mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)} \right)}{(y_t - s) h_{U_t} \left(y_t - s | \hat{\lambda}^{(k)} \right)} \mathbb{I}_{\{y_t - s > 0\}} \right. \\ \left. + \frac{\left(\sqrt{\hat{\phi}^{(k)} (\hat{\phi}^{(k)} + 2 \hat{\mu}^{(k)})} + 1 \right) (1 - \hat{\pi}^{(k)}) h_{U_t} \left(0 | \hat{\lambda}^{(k)} \right) \times \mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)} \right)}{\hat{\phi}^{(k)} \left[\hat{\pi}^{(k)} + (1 - \hat{\pi}^{(k)}) h_{U_t} \left(0 | \hat{\lambda}^{(k)} \right) \right]} \mathbb{I}_{\{y_t - s = 0\}} \right\}.$$

No passo M do algoritmo, substituímos a maximização de $\sum_{t=p+1}^n Q_t^*(\mu, \phi|\hat{\theta}^{(k)})$ pela maximização de $\sum_{t=p+1}^n Q_t^{**}(\mu, \phi|\hat{\theta}^{(k)})$. Assim, temos que a atualização de $\hat{\mu}^{(k)}$ e $\hat{\phi}^{(k)}$ é obtida por:

$$\hat{\mu}^{(k+1)} = \frac{\sum_{t=p+1}^n (1 - \widehat{w_t}^{(k)}) y_t - \sum_{t=p+1}^n \sum_{i=1}^p \widehat{b_t s_{t,i}}^{(k)}}{\sum_{t=p+1}^n (1 - \widehat{w_t}^{(k)})} \\ \hat{\phi}^{(k+1)} = \frac{\sum_{t=p+1}^n (1 - \widehat{w_t}^{(k)})}{\sum_{t=p+1}^n \widehat{b_t z_t}^{(k)} + \sum_{t=p+1}^n \widehat{b_t / z_t}^{(k)} - 2 \sum_{t=p+1}^n (1 - \widehat{w_t}^{(k)})}.$$

Em seguida, mostramos os valores iniciais que adotamos para as estimativas de α, π e λ .

Valores iniciais

Para o valor inicial $\hat{\boldsymbol{\alpha}}^{(0)}$ considerado no algoritmo EM, utilizamos o estimador baseado nas equações de Yule-Walker. Assim,

$$\hat{\boldsymbol{\alpha}}^{(0)} = R_p^{-1} \mathbf{r}_p,$$

em que $\mathbf{r}_p = (\hat{\rho}(1), \dots, \hat{\rho}(p))^\top$, sendo $\hat{\rho}(k) = \frac{1}{n-p} \sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})$ a função de autocovariância amostral, e:

$$R_p = \begin{bmatrix} 1 & \hat{\rho}(1) & \hat{\rho}(2) & \dots & \hat{\rho}(p-1) \\ \hat{\rho}(1) & 1 & \hat{\rho}(1) & \dots & \hat{\rho}(p-2) \\ \hat{\rho}(2) & \hat{\rho}(1) & 1 & \dots & \hat{\rho}(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}(p-1) & \hat{\rho}(p-2) & \hat{\rho}(p-3) & \dots & 1 \end{bmatrix}.$$

Para $\hat{\pi}^{(0)}$ e $\hat{\lambda}^{(0)}$, consideramos os estimadores de momentos igualando os momentos amostrais e populacionais. Assim, obtemos de (4.2) que:

- Se $V_t \sim \text{ZIP}(\pi, \lambda)$, então:

$$\hat{\pi}^{(0)} = \frac{\hat{\delta} \left(1 - \sum_{i=1}^p \hat{\alpha}_i^{(0)} \hat{\rho}_i \right) - 1 + \sum_{i=1}^p \left(\hat{\alpha}_i^{(0)} \right)^2}{\hat{\delta} \left(1 - \sum_{i=1}^p \hat{\alpha}_i^{(0)} \hat{\rho}_i \right) - 1 + \sum_{i=1}^p \left(\hat{\alpha}_i^{(0)} \right)^2 + \bar{y} \left(1 - \sum_{i=1}^p \hat{\alpha}_i^{(0)} \right)}$$

e

$$\hat{\lambda}^{(0)} = \frac{\bar{y} \left(1 - \sum_{i=1}^p \hat{\alpha}_i^{(0)} \right)}{1 - \hat{\pi}^{(0)}},$$

em que $\hat{\delta} = S^2 / \bar{y}$, com $S^2 = \sum_{t=1}^n (y_t - \bar{y})^2 / (n-1)$.

- Se $V_t \sim \text{ZINB}(\pi, \lambda)$ ou $V_t \sim \text{ZIPIG}(\pi, \lambda)$, então:

$$\hat{\mu}^{(0)} = \frac{\bar{y} \left(1 - \sum_{i=1}^p \hat{\alpha}_i^{(0)} \right)}{1 - \hat{\pi}^{(0)}}$$

e

$$\hat{\phi}^{(0)} = \frac{\hat{\mu}^{(0)} \left(1 - \sum_{i=1}^p \hat{\alpha}_i^{(0)} \right)}{\hat{\delta} \left(1 - \sum_{i=1}^p \hat{\alpha}_i^{(0)} \hat{\rho}_i \right) - 1 + \sum_{i=1}^p \left(\hat{\alpha}_i^{(0)} \right)^2 - \hat{\pi}^{(0)} \hat{\mu}^{(0)} \left(1 - \sum_{i=1}^p \hat{\alpha}_i^{(0)} \right)},$$

Assim como no processo ZINB-INAR(1) e ZIPIG-INAR(1), utilizamos $\hat{\pi}^{(0)} = 0,5$ como valor inicial para a estimativa de π .

4.6 PREVISÃO

Nesta seção, propomos métodos de previsão 1 passo à frente para o processo ZI-INAR(p). Assim como no processo de ordem 1, utilizamos a previsão baseada na esperança condicional, dada por:

$$\begin{aligned}\hat{y}_{n+1} &= E(Y_{n+1}|Y_n = y_n, \dots, Y_{n-p+1} = y_{n-p+1}, \hat{\boldsymbol{\theta}}) \\ &= \sum_{i=1}^p \hat{\alpha}_i y_{n-i} + (1 - \hat{\pi}) E(U_n | \hat{\boldsymbol{\lambda}}),\end{aligned}$$

em que $\hat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança obtida pelo algoritmo EM.

Como no Capítulo 3, utilizamos o EAMP(1), para avaliar a previsão 1 passo à frente no processo ZI-INAR(p), e utilizamos métodos bootstrap para obter a distribuição de previsão h passos à frente.

Assim, sendo $\mathbf{y} = (y_1, \dots, y_n)$ realizações de um processo ZI-INAR(p), a distribuição de previsão h passos à frente pode ser obtida pelos seguintes passos:

1. Escolha o método bootstrap a ser utilizado. Se o bootstrap circular em blocos for escolhido, obtenha R_1 estimativas bootstrap $\{\hat{\boldsymbol{\theta}}_r^*, r = 1, \dots, R_1\}$ como proposto para o processo ZI-INAR(1). Caso o bootstrap paramétrico seja escolhido, obtenha as R_1 estimativas bootstrap $\{\hat{\boldsymbol{\theta}}_r^*, r = 1, \dots, R_1\}$ substituindo os passos (1) e (2) do bootstrap paramétrico ZI-INAR(1) (pág. 46) por: gerar v.a.'s Y_t^* , $t = 1, 2, \dots, n$, recursivamente como:

$$\begin{aligned}Y_t^* &= y_t, \quad t = 1, \dots, p; \\ Y_t^* &= \sum_{i=1}^p \hat{\alpha}_i \circ Y_{t-i}^* + V_t^*, \quad t = p+1, \dots, n,\end{aligned}$$

em que $\hat{\alpha}_i \circ Y_{t-i}^* | Y_{t-i}^* = y_{t-i}^*$ segue uma distribuição Binomial com parâmetros y_{t-i}^* e $\hat{\alpha}_i$, se $y_{t-i}^* > 0$, e $\hat{\alpha}_i \circ Y_{t-i}^* | Y_{t-i}^* = y_{t-i}^*$ segue uma distribuição degenerada em zero, se $y_{t-i}^* = 0$, $i = 1, \dots, p$.

2. Selecione aleatoriamente uma estimativa bootstrap do conjunto $\{\hat{\boldsymbol{\theta}}_r^*, r = 1, \dots, R_1\}$ obtida no Passo (1) e gere $Y_{n+1}^*, \dots, Y_{n+h}^*$ recursivamente, utilizando a estimativa bootstrap selecionada, denotada por $\hat{\boldsymbol{\theta}}^* = (\hat{\boldsymbol{\alpha}}^*, \hat{\pi}^*, \hat{\boldsymbol{\lambda}}^*)^\top$, como segue:

$$\hat{Y}_{n+j}^* = \sum_{i=1}^p \hat{\alpha}_i^* \circ Y_{n+j-i}^* + V_{n+j}^*, \quad j = 1, \dots, h,$$

em que $Y_t^* = y_t$, para $t = n - p + 1, \dots, p$, $V_{n+1}^*, \dots, V_{n+h}^*$ são v.a.'s iid seguindo distribuição ZI e $\hat{\alpha}_i^* \circ Y_{n+j-i}^* | Y_{n+j-i}^* = y_{n+j-i}^*$ segue uma distribuição Binomial com

parâmetros y_{n+j-i}^* e $\hat{\alpha}_i^*$, se $y_{n+j-i}^* > 0$, e $\hat{\alpha}_i^* \circ Y_{n+j-i}^* | Y_{n+j-i}^* = y_{n+j-i}^*$ segue uma distribuição degenerada em zero, se $y_{n+j-i}^* = 0$, para $j = 1, \dots, h$ e $i = 1, \dots, p$.

3. Repita o Passo (2) R_2 vezes e utilize a distribuição empírica dos R_2 valores gerados para Y_{n+h}^* como a distribuição de previsão h passos à frente.

4.7 ESTUDO DE SIMULAÇÃO

Nesta seção apresentamos um estudo de simulação com o objetivo de avaliar o viés e a precisão dos estimadores de máxima verossimilhança para os parâmetros do processo ZI-INAR(p). Assim como nos estudos de simulação referente ao processo ZI-INAR(1) apresentados na Seção 3.8, utilizamos o algoritmo EM para obter as estimativas de máxima verossimilhança dos parâmetros, considerando os valores iniciais $\hat{\alpha}^{(0)}$, $\hat{\pi}^{(0)}$ e $\hat{\lambda}^{(0)}$ da subseção anterior.

Neste estudo de simulação consideramos os processos ZI-INAR de ordem $p = 2, 3$ e 4 com os seguintes parâmetros: $\pi = 0.3$ e 0.6 e, $\alpha = (\alpha_1, \alpha_2) = (0.3, 0.2)$, para o processo ZI-INAR(2), $\alpha = (\alpha_1, \alpha_2, \alpha_3) = (0.25, 0.2, 0.15)$ para o processo ZI-INAR(3) e $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.25, 0.2, 0.15, 0.1)$ para o processo ZI-INAR(4). Considerando os diferentes valores de p , geramos $N = 300$ réplicas dos processos: ZIP-INAR(p) com parâmetro $\lambda = 2$; ZINB-INAR(p) e ZIPIG-INAR(p) com parâmetros $\mu = 2$ e $\phi = 1.5$. Os tamanhos de amostras adotados foram $n = 100, 300, 500$ e 1000 .

Conforme observado nas Tabelas 13-15, o viés relativo (VR) das estimativas dos parâmetros dos processos ZI-INAR considerados no estudo estão em torno de zero, enquanto que a raiz quadrada do erro quadrático médio relativo (REQMR) diminui em todos os cenários, quando o tamanho da amostra aumenta. Nota-se ainda que nos processos com inovações ZINB e ZIPIG as estimativas do parâmetro ϕ apresentaram maiores VR e REQMR quando comparadas com as estimativas dos demais parâmetros.

Sendo assim, concluímos que as estimativas dos parâmetros do processo se comportaram como o esperado, ou seja, se apresentaram em torno do verdadeiro valor do parâmetro, e apresentaram maior precisão para tamanhos maiores da amostra.

Tabela 13 – VR e REQMR das estimativas dos parâmetros do processo ZIP-INAR(p) para diferentes tamanhos de amostra, fixando $\pi = 0.3$ e 0.6 e $\lambda = 2$, considerando as ordens $p = 2$, com $\alpha = (0.3, 0.2)$, $p = 3$, com $\alpha = (0.25, 0.2, 0.15)$ e $p = 4$, com $\alpha = (0.25, 0.2, 0.15, 0.1)$.

p	π	n	$\hat{\alpha}_1$		$\hat{\alpha}_2$		$\hat{\alpha}_3$		$\hat{\alpha}_4$		$\hat{\pi}$		$\hat{\lambda}$	
			VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR
2	0.3	100	-0.060	0.428	-0.106	0.399	-	-	-	-	-0.142	0.476	0.037	0.176
		300	-0.003	0.240	-0.058	0.252	-	-	-	-	-0.054	0.279	0.013	0.096
		500	-0.004	0.173	-0.018	0.183	-	-	-	-	-0.026	0.210	-0.000	0.070
		1000	-0.010	0.120	-0.005	0.138	-	-	-	-	-0.017	0.150	0.006	0.053
	0.6	100	-0.020	0.332	-0.103	0.398	-	-	-	-	-0.056	0.178	-0.014	0.158
		300	-0.004	0.195	-0.032	0.203	-	-	-	-	-0.018	0.097	-0.010	0.101
		500	0.004	0.146	-0.029	0.157	-	-	-	-	-0.009	0.076	-0.003	0.084
		1000	-0.009	0.105	-0.017	0.115	-	-	-	-	-0.009	0.048	-0.006	0.052
3	0.3	100	-0.098	0.580	-0.128	0.506	-0.001	0.529	-	-	-0.159	0.612	0.057	0.218
		300	-0.020	0.291	-0.010	0.305	-0.041	0.289	-	-	-0.009	0.349	0.032	0.123
		500	-0.012	0.242	-0.014	0.215	-0.044	0.237	-	-	-0.046	0.271	0.010	0.092
		1000	-0.010	0.168	-0.002	0.166	-0.015	0.151	-	-	-0.019	0.192	0.007	0.059
	0.6	100	-0.019	0.506	-0.099	0.522	-0.055	0.469	-	-	-0.094	0.281	-0.016	0.215
		300	-0.002	0.263	-0.043	0.292	-0.037	0.285	-	-	-0.020	0.121	0.008	0.120
		500	0.003	0.198	-0.024	0.216	-0.035	0.216	-	-	-0.018	0.095	-0.001	0.095
		1000	0.008	0.150	-0.029	0.144	-0.014	0.151	-	-	-0.010	0.067	-0.004	0.063
4	0.3	100	-0.044	0.730	-0.180	0.721	-0.063	0.674	-0.104	0.602	-0.219	0.701	0.134	0.345
		300	-0.017	0.369	-0.041	0.384	-0.013	0.410	-0.150	0.381	-0.131	0.417	0.041	0.164
		500	0.006	0.262	-0.036	0.297	0.001	0.328	-0.078	0.332	-0.082	0.342	0.016	0.112
		1000	-0.008	0.198	-0.011	0.215	0.004	0.222	-0.011	0.218	-0.027	0.251	0.005	0.077
	0.6	100	-0.053	0.679	-0.131	0.667	-0.039	0.628	-0.097	0.556	-0.161	0.400	0.024	0.279
		300	-0.006	0.372	-0.046	0.405	-0.010	0.368	-0.048	0.366	-0.050	0.184	-0.010	0.141
		500	0.003	0.289	-0.034	0.253	-0.018	0.288	-0.021	0.290	-0.032	0.128	-0.006	0.112
		1000	0.001	0.199	-0.009	0.217	0.003	0.213	-0.043	0.194	-0.010	0.090	0.004	0.080

Tabela 14 – VR e REQMR das estimativas dos parâmetros do processo ZINB-INAR(p) para diferentes tamanhos de amostra, fixando $\pi = 0.3$ e 0.6 , $\mu = 2$, e $\phi = 1.5$, considerando as ordens $p = 2$, com $\alpha = (0.3, 0.2)$, $p = 3$, com $\alpha = (0.25, 0.2, 0.15)$ e $p = 4$, com $\alpha = (0.25, 0.2, 0.15, 0.1)$.

p	π	n	$\hat{\alpha}_1$		$\hat{\alpha}_2$		$\hat{\alpha}_3$		$\hat{\alpha}_4$		$\hat{\pi}$		$\hat{\mu}$		$\hat{\phi}$	
			VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR
2	0.3	100	-0.014	0.340	-0.101	0.341	-	-	-	-	-0.186	0.728	0.054	0.324	0.744	2.052
		300	-0.002	0.195	-0.024	0.207	-	-	-	-	-0.082	0.533	0.020	0.203	0.345	1.230
		500	-0.007	0.162	-0.017	0.162	-	-	-	-	-0.081	0.462	0.009	0.173	0.234	0.764
		1000	0.002	0.106	-0.008	0.110	-	-	-	-	-0.037	0.341	-0.002	0.122	0.084	0.386
	0.6	100	0.003	0.314	-0.087	0.330	-	-	-	-	-0.162	0.441	0.009	0.398	0.878	2.219
		300	-0.007	0.177	-0.027	0.166	-	-	-	-	-0.082	0.266	-0.041	0.244	0.329	1.271
		500	0.007	0.136	-0.036	0.125	-	-	-	-	-0.041	0.187	0.007	0.197	0.300	1.028
		1000	0.001	0.101	0.007	0.098	-	-	-	-	-0.014	0.107	-0.004	0.137	0.101	0.452
3	0.3	100	0.001	0.450	-0.097	0.484	-0.101	0.445	-	-	-0.209	0.791	0.086	0.374	1.064	2.346
		300	-0.004	0.269	-0.045	0.251	-0.026	0.272	-	-	-0.174	0.646	0.002	0.229	0.481	1.401
		500	0.004	0.197	-0.005	0.194	-0.010	0.199	-	-	-0.097	0.564	0.008	0.206	0.286	0.995
		1000	0.004	0.147	-0.012	0.153	-0.012	0.156	-	-	-0.050	0.420	0.012	0.149	0.140	0.534
	0.6	100	-0.008	0.426	-0.024	0.448	-0.111	0.434	-	-	-0.314	0.578	-0.090	0.446	0.530	1.896
		300	0.007	0.230	-0.028	0.219	-0.011	0.230	-	-	-0.063	0.287	0.016	0.282	0.581	1.570
		500	0.006	0.180	-0.032	0.184	-0.032	0.183	-	-	-0.075	0.267	-0.006	0.241	0.387	1.176
		1000	-0.009	0.135	-0.005	0.135	0.0020	0.126	-	-	-0.036	0.173	-0.007	0.186	0.165	0.672
4	0.3	100	-0.034	0.602	-0.081	0.606	-0.119	0.620	-0.142	0.514	-0.346	0.812	0.091	0.369	1.166	2.503
		300	-0.014	0.335	-0.025	0.317	-0.028	0.384	0.012	0.340	-0.134	0.731	0.042	0.284	0.697	1.762
		500	-0.011	0.249	-0.016	0.296	-0.008	0.318	-0.024	0.267	-0.099	0.651	0.040	0.239	0.635	1.754
		1000	-0.002	0.181	-0.004	0.196	-0.008	0.206	-0.026	0.204	-0.114	0.535	0.002	0.179	0.198	0.757
	0.6	100	-0.023	0.606	-0.130	0.614	-0.056	0.530	-0.052	0.513	-0.442	0.672	-0.117	0.487	0.368	1.775
		300	-0.008	0.345	-0.004	0.338	-0.029	0.353	-0.056	0.308	-0.151	0.412	-0.017	0.346	0.726	1.965
		500	0.001	0.250	-0.014	0.266	-0.004	0.263	-0.031	0.262	-0.094	0.332	-0.010	0.285	0.422	1.252
		1000	0.001	0.154	-0.020	0.175	-0.004	0.192	-0.010	0.169	-0.065	0.235	-0.025	0.219	0.206	0.802

Tabela 15 – VR e REQMR das estimativas dos parâmetros do processo ZIPIG-INAR(p) para diferentes tamanhos de amostra, fixando $\pi = 0.3$ e 0.6 , $\mu = 2$, e $\phi = 1.5$, considerando as ordens $p = 2$, com $\alpha = (0.3, 0.2)$, $p = 3$, com $\alpha = (0.25, 0.2, 0.15)$ e $p = 4$, com $\alpha = (0.25, 0.2, 0.15, 0.1)$.

p	π	n	$\hat{\alpha}_1$		$\hat{\alpha}_2$		$\hat{\alpha}_3$		$\hat{\alpha}_4$		$\hat{\pi}$		$\hat{\mu}$		$\hat{\phi}$	
			VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR	VR	RE-QMR
2	0.3	100	-0.019	0.335	-0.095	0.359	-	-	-	-	-0.190	0.666	0.026	0.268	0.583	1.576
		300	-0.018	0.212	-0.017	0.183	-	-	-	-	-0.106	0.484	0.006	0.161	0.281	0.977
		500	-0.009	0.142	-0.000	0.167	-	-	-	-	-0.025	0.348	0.008	0.122	0.173	0.589
		1000	0.000	0.104	-0.015	0.113	-	-	-	-	-0.029	0.244	0.002	0.093	0.058	0.308
	0.6	100	-0.020	0.325	-0.054	0.318	-	-	-	-	-0.112	0.351	0.012	0.358	0.660	1.644
		300	0.006	0.172	-0.031	0.172	-	-	-	-	-0.022	0.166	0.011	0.195	0.410	1.168
		500	-0.002	0.130	-0.025	0.136	-	-	-	-	-0.011	0.116	0.002	0.158	0.245	0.749
		1000	0.003	0.092	-0.005	0.094	-	-	-	-	0.003	0.072	0.012	0.105	0.108	0.399
3	0.3	100	-0.052	0.493	-0.101	0.499	-0.133	0.407	-	-	-0.280	0.760	0.050	0.316	0.927	2.152
		300	-0.041	0.284	-0.025	0.257	0.002	0.283	-	-	-0.148	0.608	0.012	0.210	0.351	1.070
		500	-0.024	0.194	-0.013	0.221	0.003	0.203	-	-	-0.065	0.452	0.012	0.148	0.231	0.680
		1000	0.005	0.143	-0.027	0.142	-0.015	0.143	-	-	-0.036	0.347	0.012	0.115	0.127	0.424
	0.6	100	-0.064	0.412	-0.100	0.436	0.010	0.432	-	-	-0.207	0.474	-0.022	0.406	0.734	2.055
		300	0.014	0.222	-0.008	0.248	-0.037	0.244	-	-	-0.031	0.226	0.035	0.239	0.533	1.335
		500	0.0002	0.187	-0.023	0.195	-0.007	0.181	-	-	-0.036	0.171	-0.005	0.190	0.204	0.762
		1000	0.005	0.131	-0.010	0.132	-0.008	0.142	-	-	-0.010	0.107	-0.001	0.134	0.102	0.412
4	0.3	100	-0.040	0.669	-0.081	0.680	-0.127	0.645	-0.134	0.486	-0.424	0.864	0.070	0.370	0.951	2.199
		300	-0.018	0.381	-0.008	0.359	-0.024	0.365	-0.118	0.331	-0.153	0.662	0.059	0.259	0.513	1.447
		500	0.002	0.280	-0.035	0.285	-0.024	0.289	-0.013	0.303	-0.136	0.539	0.012	0.196	0.269	0.905
		1000	-0.009	0.190	-0.009	0.192	0.006	0.189	-0.049	0.185	-0.060	0.455	0.022	0.148	0.194	0.599
	0.6	100	-0.034	0.565	-0.114	0.626	-0.060	0.604	-0.160	0.521	-0.388	0.613	-0.099	0.438	0.633	2.028
		300	-0.002	0.335	-0.036	0.340	-0.013	0.314	-0.041	0.344	-0.135	0.357	-0.040	0.287	0.474	1.479
		500	0.002	0.265	-0.046	0.276	0.010	0.261	-0.029	0.253	-0.056	0.246	0.010	0.236	0.371	1.192
		1000	-0.008	0.173	0.007	0.170	-0.017	0.193	-0.002	0.168	-0.031	0.176	0.0022	0.175	0.153	0.567

4.8 APLICAÇÃO

Nesta seção, será apresentada uma aplicação a dados reais, utilizando os processos ZI-INAR(p) e as metodologias propostas neste capítulo.

O conjunto de dados se refere ao número mensal de crimes ocorridos no distrito 1303 de São Petersburgo no período de janeiro de 1990 até dezembro de 2001. Este conjunto de dados pode ser encontrado na mesma base de dados utilizada para a aplicação do processo ZI-INAR(1), disponível no site (<http://www.forecastingprinciples.com>).

As estatísticas descritivas da série são apresentadas na Tabela 16. Conforme observamos, o conjunto de dados apresenta 52.8% de zeros e uma variância maior que a média.

Tabela 16 – Estatísticas descritivas da série *número de crimes no distrito 1303*.

n	144
Mínimo	0
Média	1.576
Mediana	0
Máximo	14
Variância	6.610
Proporção de zeros	0.528

A Figura 4 apresenta o gráfico da série, o gráfico de barras e a função de autocorrelação parcial (FACP). Analisando o gráfico da FACP, temos indícios que um

modelo autorregressivo de ordem maior que 1 seja adequado.

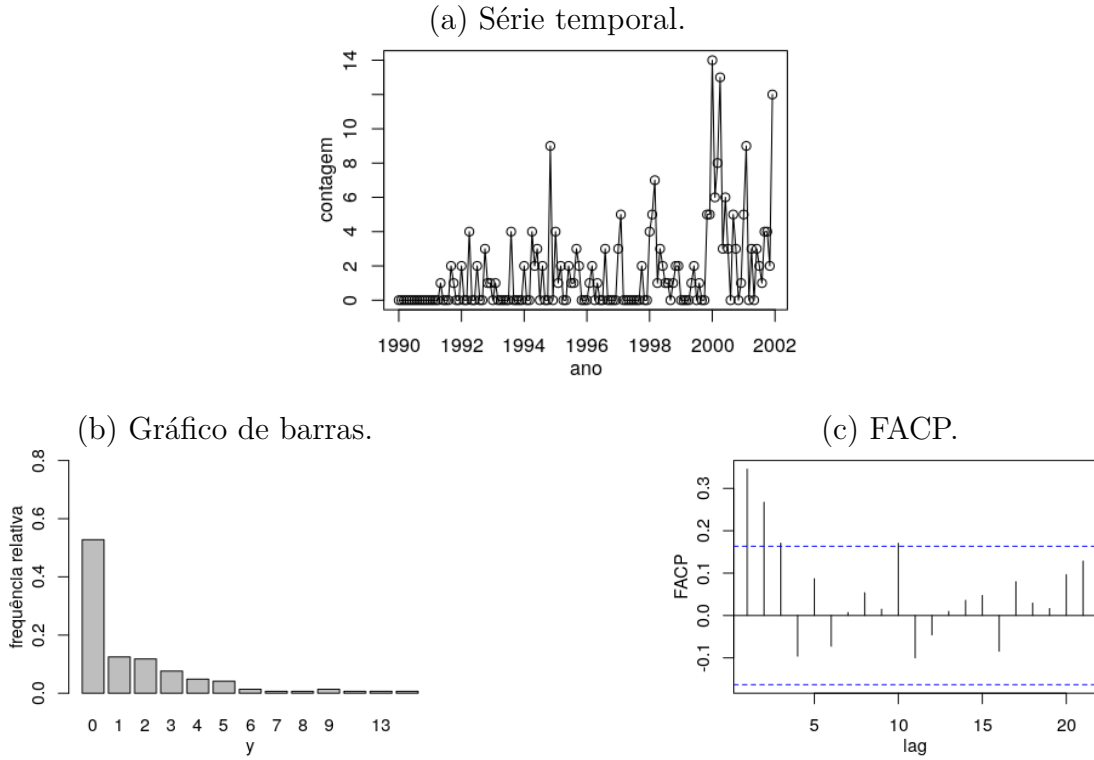


Figura 4 – Série número de crimes no distrito 1303.

Estimação:

A Tabela 17 apresenta as estimativas dos parâmetros dos processos Po-INAR, ZIP-INAR, NB-INAR, ZINB-INAR, PIG-INAR e ZIPIG-INAR para as ordens $p = 1, 2$ e 3 , assim como as medidas do AIC. Para comparar modelos de diferentes ordens, utilizamos no cálculo do AIC a função de verossimilhança condicionada às observações $y_1, \dots, y_{p_{max}}$, em que p_{max} é a maior ordem dos modelos comparados. Assim, o AIC é obtido por:

$$AIC = -2 \sum_{t=p_{max}+1}^n \log \left[P(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}) \right] + 2k,$$

em que k é o número de parâmetros do modelo. Note que na nossa aplicação, utilizamos $p_{max} = 3$. Conforme observamos, a medida do AIC sugere que um processo de ordem $p > 1$ seja adequado e que o processo de segunda ordem com inovações ZIPIG seja utilizado em comparação aos demais.

A Figura 5 apresenta as distribuições bootstrap dos estimadores dos parâmetros do processo ZIPIG-INAR(2), considerando $R = 100$ réplicas de bootstrap para os dois métodos propostos na Seção 3.6. Como sugerido por Bisaglia e Gerolimetto (2019), para o método por reamostragem em blocos, consideramos tamanhos de blocos $l = \sqrt{n}$, i.e.,

Tabela 17 – Estimativa dos parâmetros para os processos INAR(p) com inovações Poisson, ZIP, NB, ZINB, PIG e ZIPIG. Série: *crimes no distrito 1303*

Processo	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\pi}$	$\hat{\mu}$	$\hat{\phi}$	AIC
Po-INAR(1)	0.234				1.236		622.43
Po-INAR(2)	0.197	0.175			1.037		602.07
Po-INAR(3)	0.175	0.140	0.151		0.908		591.36
ZIP-INAR(1)	0.208			0.603	3.212		496.08
ZIP-INAR(2)	0.141	0.198		0.674	3.335		482.55
ZIP-INAR(3)	0.122	0.162	0.072	0.674	3.292		483.14
NB-INAR(1)	0.128				1.395	0.364	476.67
NB-INAR(2)	0.093	0.153			1.228	0.302	474.08
NB-INAR(3)	0.066	0.123	0.076		1.210	0.301	474.42
ZINB-INAR(1)	0.135			0.450	2.519	1.333	475.63
ZINB-INAR(2)	0.096	0.162		0.557	2.733	1.763	471.59
ZINB-INAR(3)	0.076	0.132	0.065	0.551	2.668	1.657	472.50
PIG-INAR(1)	0.126				1.397	0.265	484.99
PIG-INAR(2)	0.098	0.148			1.229	0.222	482.59
PIG-INAR(3)	0.065	0.119	0.087		1.202	0.218	482.34
ZIPIG-INAR(1)	0.132			0.496	2.757	1.540	474.89
ZIPIG-INAR(2)	0.093	0.160		0.579	2.893	1.900	470.75
ZIPIG-INAR(3)	0.074	0.132	0.062	0.577	2.846	1.823	471.75

$l = 12$. É possível perceber que, assim como na aplicação para o processo de primeira ordem apresentada na Seção 3.9, observamos diversas amostras bootstrap cujas estimativas de ϕ se distanciaram da média, especialmente para o bootstrap paramétrico.

Na Tabela 18 são apresentadas as estimativas dos parâmetros e os intervalos de confiança obtidos via bootstrap paramétrico e via bootstrap circular em blocos. Note que, comparando os dois métodos bootstrap, observamos que as estimativas do método paramétrico se apresentaram mais próximos das estimativas do algoritmo EM, com exceção da estimativa de ϕ . Neste caso, os métodos não forneceram valores de $\hat{\phi}$ próximo ao que foi observado no algoritmo EM.

Tabela 18 – Estimativas e intervalos de confiança para os parâmetros do processo ZIPIG-INAR(2). B-PAR (Bootstrap paramétrico). BCB (Bootstrap circular em blocos).

Parâmetro	Estimativa			Desv. Pad.		Intervalo bootstrap	
	EM	B-PAR	BCB	B-PAR	BCB	B-PAR	BCB
α_1	0.093	0.098	0.058	0.044	0.053	[0.013,0.184]	[0,0.161]
α_2	0.160	0.147	0.077	0.045	0.061	[0.059,0.234]	[0,0.195]
π	0.579	0.575	0.512	0.091	0.088	[0.397,0.752]	[0.340,0.685]
μ	2.893	2.868	2.797	0.563	0.599	[1.764,3.972]	[1.624,3.970]
ϕ	1.900	7.850	2.333	21.936	2.091	(0,50.843]	(0,6.431]

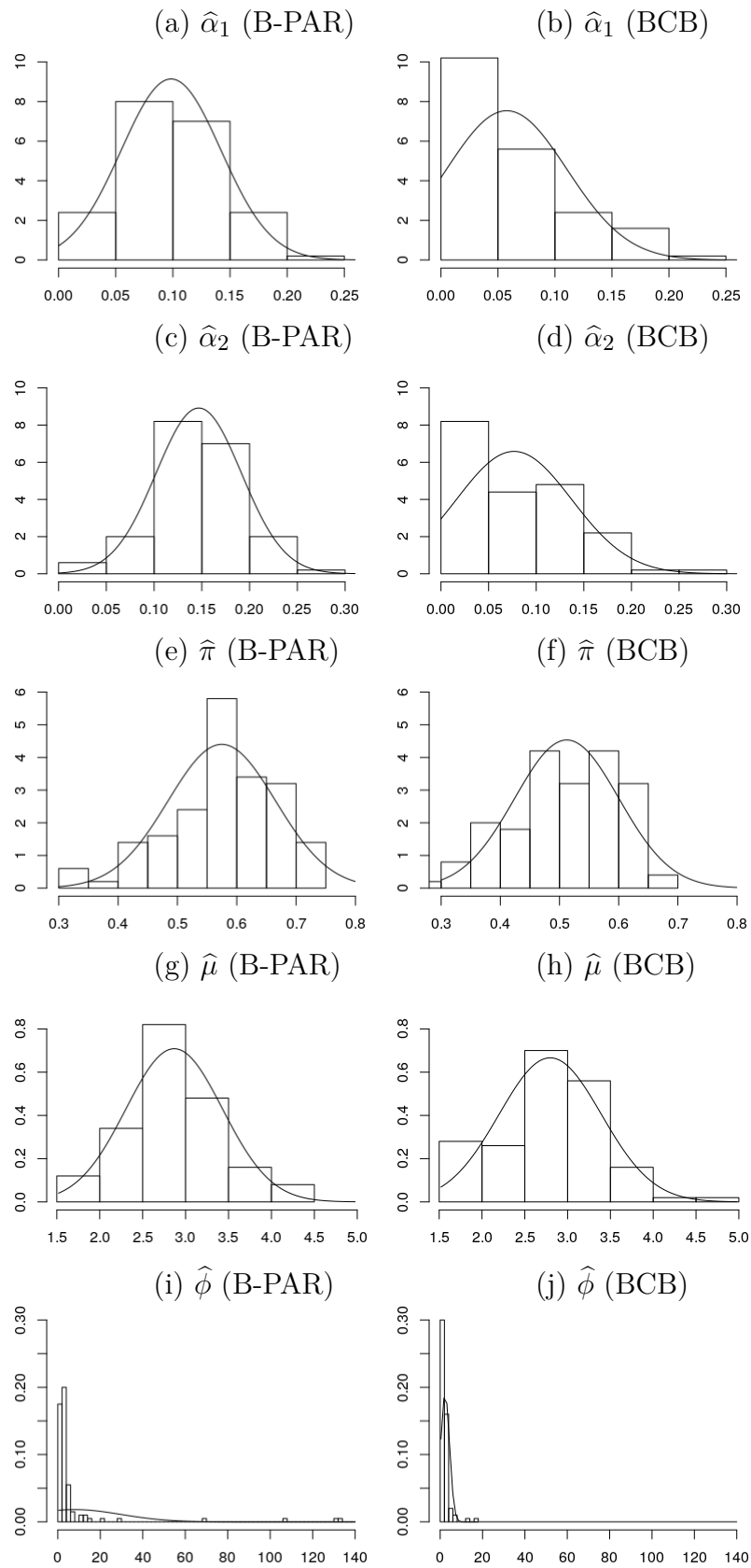


Figura 5 – Distribuição dos estimadores dos parâmetros do processo ZIPIG-INAR(2) obtidas pelos métodos de bootstrap paramétrico (B-PAR) e bootstrap circular em blocos (BCB).

Previsão:

A Tabela 19 apresenta o EAMP(1) (veja Eq. (3.40) – Cap. 3), para o qual utilizamos as primeiras 132 observações da série para estimação dos parâmetros, e as $m = 12$ últimas para previsões. É possível observar que quando comparamos processos com mesma distribuição para a inovação, os processos de maior ordem, fornecem melhores previsões, de acordo com a medida adotada. Além disso, quando comparamos processos de mesma ordem, os processos com inovações seguindo distribuição ZIP, ZINB e ZIPIG apresentam menores valores de EAMP(1) que os processos com inovações seguindo distribuição Poisson, Binomial Negativa e Poisson Inversa Gaussiana, respectivamente.

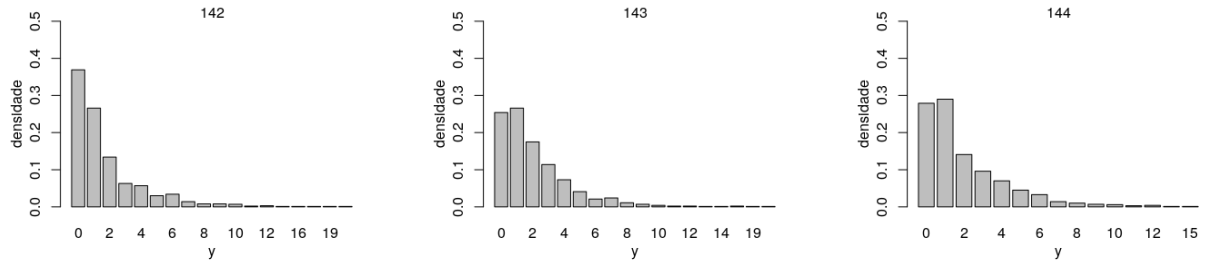
Tabela 19 – EAMP(1) das $m = 12$ últimas observações da série *número de crimes no distrito 1303*.

Processo	p		
	1	2	3
Po-INAR	2.940	2.891	2.854
ZIP-INAR	2.907	2.843	2.833
NB-INAR	2.888	2.829	2.821
ZINB-INAR	2.889	2.826	2.817
PIG-INAR	2.890	2.834	2.819
ZIPIG-INAR	2.889	2.824	2.817

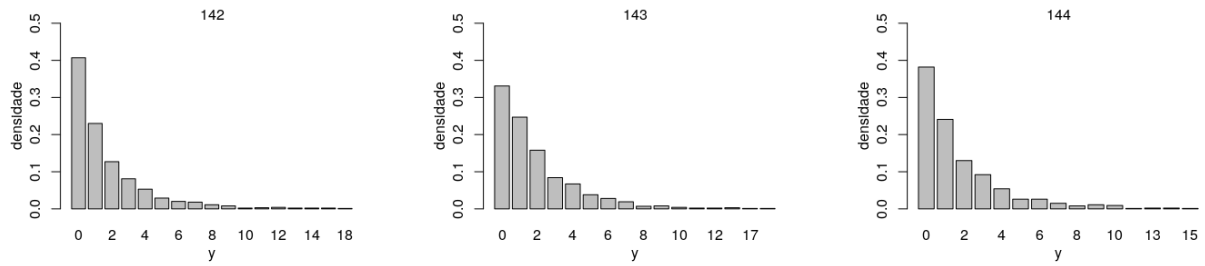
Comparando os processos pelos valores do AIC e pelas medidas de previsão, sugerimos utilizar o processo ZIPIG-INAR(2) para modelar a série, uma vez que este apresentou menor valor do AIC, e suas previsões forneceram melhores resultados de acordo com o EAMP(1), quando comparado com os demais processos de segunda ordem.

A Figura 3 apresenta as distribuições de previsão $h = 1, 2$ e 3 passos à frente das três últimas observações da série (y_{142} , y_{143} e y_{144}) obtidas pelo ajuste do processo ZIPIG-INAR(2), i.e., a distribuição de $Y_{142}|Y_{141} = y_{141}$, $Y_{143}|Y_{141} = y_{141}$ e $Y_{144}|Y_{141} = y_{141}$. Para obter estas distribuições de previsão seguimos os passos (1)-(3) dados na Seção 4.6. No passo (1), geramos $R_1 = 100$ estimativas bootstrap, utilizando as observações $y_1, \dots, y_{141}$ da série, e nos passos (2) e (3), geramos $R_2 = 1000$ valores para y_{142}^* , y_{143}^* e y_{144}^* . É possível perceber que as distribuições de previsão obtidas pelo método de bootstrap paramétrico apresentam comportamentos similares às obtidas pelo método de bootstrap em blocos circulares. Em geral, observamos um comportamento decrescente destas distribuições, tendo o zero como o valor modal. Utilizando o método de bootstrap

paramétrico, estima-se que as probabilidades de observar os três últimos valores da série ($y_{142} = 4$, $y_{143} = 2$ e $y_{144} = 12$), dado $y_1, \dots, y_{141}$, são 0,057, 0,175 e 0,004, respectivamente, e utilizando o método de reamostragem em blocos circulares, estas probabilidades são 0,053, 0,158 e 0,002.



(a) Bootstrap paramétrico.



(b) Bootstrap circular em blocos.

Figura 6 – Distribuição de previsão $h = 1, 2$ e 3 passos à frente para a série temporal *número de crimes no distrito 1303*.

5 CONCLUSÕES E TRABALHOS FUTUROS

5.1 CONCLUSÕES

Nesta tese, introduzimos uma nova classe de processos INAR, com inovações seguindo distribuições Inflacionadas de Zeros, denotada por ZI-INAR, uma generalização do processo ZINAR(1), proposto por Jazi, Jones e Lai (2012). Para a classe, desenvolvemos um algoritmo EM para estimar os parâmetros, e propomos intervalos de confiança bootstrap, além de métodos de previsão.

Realizamos os estudos do processo ZI-INAR de primeira ordem (ZI-INAR(1)) e estudamos o processo de ordem p (ZI-INAR(p)), considerando os casos particulares em que as inovações seguem distribuição ZIP, ZINB e ZIPIG. Os estudos de simulação mostraram que o viés relativo das estimativas dos parâmetros se concentram em torno de zero, e a raiz quadrada do erro quadrático médio decresce quando o tamanho da amostra aumenta. Aplicações a dados reais mostraram que os processos com inovações seguindo distribuição ZI se ajustaram melhor à série (de acordo com a medida do AIC) e apresentaram melhores previsões (de acordo com o erro absoluto médio de previsão) quando comparado com os processos cujas inovações não seguem distribuição ZI.

Para desenvolver o algoritmo EM para o processo ZI-INAR(p), propomos um teorema que pode ser útil no problema de estimação em processos INAR. Este teorema propõe uma transformação linear que contorna o problema de maximizar a verossimilhança sob a restrição de que a soma dos elementos do vetor autorregressivo seja menor que 1 (condições para as quais o processo é estacionário).

5.2 TRABALHOS FUTUROS

Considerando os resultados obtidos nesta tese, diversas pesquisas podem ser realizadas futuramente. Dentre estas pesquisas, destacamos as seguintes:

1. Estender os resultados de estimação e previsão obtidos no processo ZI-INAR(p) para o processo autorregressivo de médias móveis para valores inteiros não negativos, INARMA(p, q), considerando inovações seguindo distribuições ZI.
2. Introduzir novos processos da classe ZI-INAR, como por exemplo, o processo INAR com inovações seguindo uma distribuição Binomial Inflacionada de Zeros e Beta-Binomial Inflacionada de zeros.

3. Desenvolver pacotes estatísticos que possam ser instalados e utilizados de forma gratuita, possibilitando o usuário a realizar estimação e previsão do processos ZI-INAR.

REFERÊNCIAS

- ABRAMOWITZ, M.; STEGUN, I. Handbook of mathematical functions: With formulas, graphs, and mathematical tables applied mathematics series. **National Bureau of Standards, Washington, DC**, 1964.
- AL-OSH, M.; ALZAID, A. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, v. 8, p. 261–275, 1987.
- ALZAID, A.; AL-OSH, M. First-order integer-valued autoregressive (INAR (1)) process: distributional and regression properties. *Statistica Neerlandica*, v. 42, p. 53–61, 1988.
- BARRETO-SOUZA, W. Zero-Modified Geometric INAR(1) process for modelling count time series with deflation or inflation of zeros. *Journal of Time Series Analysis*, v. 36, p. 839–852, 2015.
- BARRETO-SOUZA, W.; SIMAS, A. General mixed poisson regression models with varying dispersion. *Statistics and Computing*, v. 26, p. 1263–1280, 2016.
- BISAGLIA, L.; GEROLIMETTO, M. Model-based INAR bootstrap for forecasting INAR(p) models. *Computational Statistics*, v. 34, p. 1815–1848, 2019.
- BU, R.; MCCABE, B. Model selection, estimation and forecasting in INAR(p) models: A likelihood-based markov chain approach. *International Journal of Forecasting*, v. 24, p. 151–162, 2008.
- BU, R.; MCCABE, B.; KADDOUR, H. Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of Time Series Analysis*, v. 29, p. 973–994, 2008.
- CARDINAL, M.; ROY, R.; LAMBERT, J. On the application of integer-valued time series models for the analysis of disease incidence. *Statistics in Medicine*, v. 18, p. 2025–2039, 1999.
- DAVISON, A.; HINKLEY, D. **Bootstrap methods and their application**. [S.l.]: Cambridge university press, 1997.
- DEMPSTER, A.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B*, v. 39, p. 1–38, 1977.
- DION, J.; GAUTHIER, G.; LATOUR, A. Branching processes with immigration and integer-valued time series. *Serdica Mathematical Journal*, v. 21, p. 123–136, 1995.
- DU, J.; LI, Y. The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, v. 12, p. 129–142, 1991.
- EFRON, B. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, v. 7, p. 1–16, 1979.
- FELLER, W. **An introduction to probability theory and its applications**. [S.l.]: John Wiley & Sons, 2008. v. 2.
- FRANKE, J.; RAO, T. **Multivariate First-order Integer-valued Autoregressions**. [S.l.]: Arbeitsgruppe Technomathematik, Univ., 1993.

- GARAY, A.; BOLFARINE, H.; V.H, L.; R.B.C., C. Bayesian analysis of censored linear regression models with scale mixtures of normal distributions. *Journal of Applied Statistics*, v. 42, p. 2694–2714, 2015.
- GARAY, A.; ELIZABETH, M.; EDWIN, M. O.; LACHOS, V. On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics and Data Analysis*, v. 55, p. 1304–1318, 2011.
- HALL, D. Zero-inflated Poisson and Binomial regression with random effects: a case study. *Biometrics*, v. 56, p. 1030–1039, 2000.
- JAMES, B. **Probabilidade: um curso em nível intermediário**. [S.l.]: IMPA, Rio de Janeiro, 2010.
- JAZI, M.; JONES, G.; LAI, C. First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis*, v. 33, p. 954–963, 2012.
- JENTSCH, C.; WEISS, C. Bootstrapping INAR models. *Bernoulli*, v. 25, p. 2359–2408, 2019.
- JOHNSON, N.; KEMP, A.; KOTZ, S. **Univariate discrete distributions**. [S.l.]: John Wiley & Sons, 2005.
- JORGENSEN, B. **Statistical properties of the generalized inverse Gaussian distribution**. [S.l.]: Springer Science & Business Media, 2012. v. 9.
- JUNG, R.; TREMAYNE, A. Coherent forecasting in integer time series models. *International Journal of Forecasting*, v. 22, p. 223–238, 2006.
- KARLIS, D. A general em approach for maximum likelihood estimation in mixed poisson regression models. *Statistical Modelling*, v. 1, p. 305–318, 2001.
- KIM, H.; PARK, Y. A non-stationary integer-valued autoregressive model. *Statistical Papers*, v. 49, p. 485, 2008.
- KIM, H.; PARK, Y. Coherent forecasting in binomial AR(p) model. *Communications for Statistical Applications and Methods*, v. 17, p. 27–37, 2010.
- KUNSCH, H. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, v. 17, p. 1217–1241, 1989.
- LAMBERT, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, v. 34, p. 1–14, 1992.
- LATOURE, A. Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, v. 19, p. 439–455, 1998.
- LIU, C.; RUBIN, D. The ecme algorithm: a simple extension of em and ecm with faster monotone convergence. *Biometrika*, v. 81, p. 633–648, 1994.
- MAITI, R.; BISWAS, A.; GUHA, A.; ONG, S. Modelling and coherent forecasting of zero-inflated count time series. *Statistical Modelling*, v. 14, p. 375–398, 2014.
- MARTINEZ-FLORES, G.; BOLFARINE, H.; GOMEZ, H. Asymmetric regression models with limited responses with an application to antibody response to vaccine. *Biometrical Journal*, v. 55, p. 156–172, 2013.

- MCKENZIE. Some simple models for discrete variate time series. **JAWRA Journal of the American Water Resources Association**, v. 21, p. 645–650, 1985.
- MCLACHLAN, G.; KRISHNAN, T. **The EM Algorithm and Extensions**. second edition. [S.l.]: John Wiley and Sons, New York,, 2008.
- MENG, X.; RUBIN, D. Maximum likelihood estimation via the ecm algorithm: A general framework. **Biometrika**, v. 80, p. 267–278, 1993.
- NEAL, P.; RAO, T. S. MCMC for integer-valued ARMA processes. **Journal of Time Series Analysis**, v. 28, p. 92–110, 2007.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<http://www.R-project.org/>>.
- RIDOUT, M.; HINDE, J.; DEMÉTRIO, C. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. **Biometrics**, v. 57, p. 219–223, 2001.
- RISTIĆ, M.; BAKOUCH, H.; NASTIĆ, A. A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. **Journal of Statistical Planning and Inference**, v. 139, p. 2218–2226, 2009.
- RISTIĆ, M. M.; BOURGUIGNON, M.; NASTIĆ, A. S. Zero-inflated NGINAR(1) process. **Communications in Statistics-Theory and Methods**, v. 48, p. 726–741, 2019.
- SAPATINAS, T. Identifiability of mixtures of power-series distributions and related characterizations. **Annals of the Institute of Statistical Mathematics**, v. 47, p. 447–459, 1995.
- SILVA, I.; SILVA, M. Asymptotic distribution of the yule–walker estimator for INAR(p) processes. **Statistics & probability letters**, v. 76, p. 1655–1663, 2006.
- SILVA, M.; OLIVEIRA, V. Difference equations for the higher-order moments and cumulants of the INAR(1) model. **Journal of Time Series Analysis**, v. 25, p. 317–333, 2004.
- STEUTEL, F. W.; HARN, K. V. Discrete analogues of self-decomposability and stability. **The Annals of Probability**, v. 7, p. 893–899, 1979.
- TAY, A.; WALLIS, K. Density forecasting: a survey. **Journal of forecasting**, v. 19, p. 235–254, 2000.
- VAIDA, F. Parameter convergence for EM and MM algorithms. **Statistica Sinica**, v. 15, p. 831–840, 2005.
- WU, C. On the convergence properties of the EM algorithm. **The Annals of Statistics**, v. 11, p. 95–103, 1983.
- ZELLER, C.; CABRAL, C.; LACHOS, V.; BENITES, L. Finite mixture of regression models for censored data based on scale mixtures of normal distributions. **Advances in Data Analysis and Classification**, v. 13, p. 89–116, 2019.

APÊNDICE A – RESULTADOS DO ESTUDO DE SIMULAÇÃO 2 DO CAPÍTULO 3

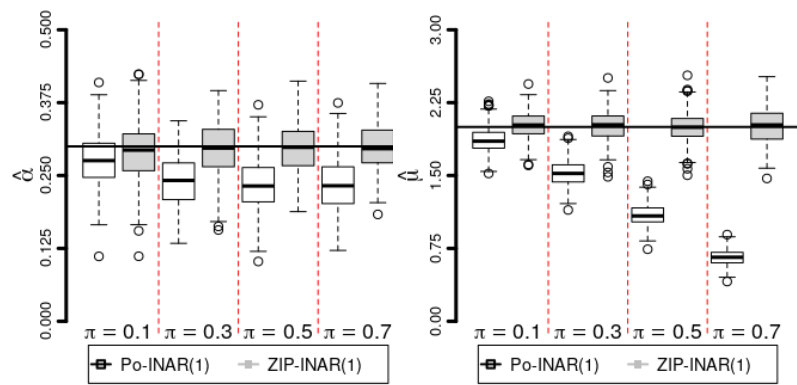


Figura 7 – Box-plot das estimativas dos parâmetros para os processos Po-INAR(1) e ZIP-INAR(1) considerando $\mu = 2$ e $n = 300$. Linha horizontal contínua preta: verdadeiro valor do parâmetro.

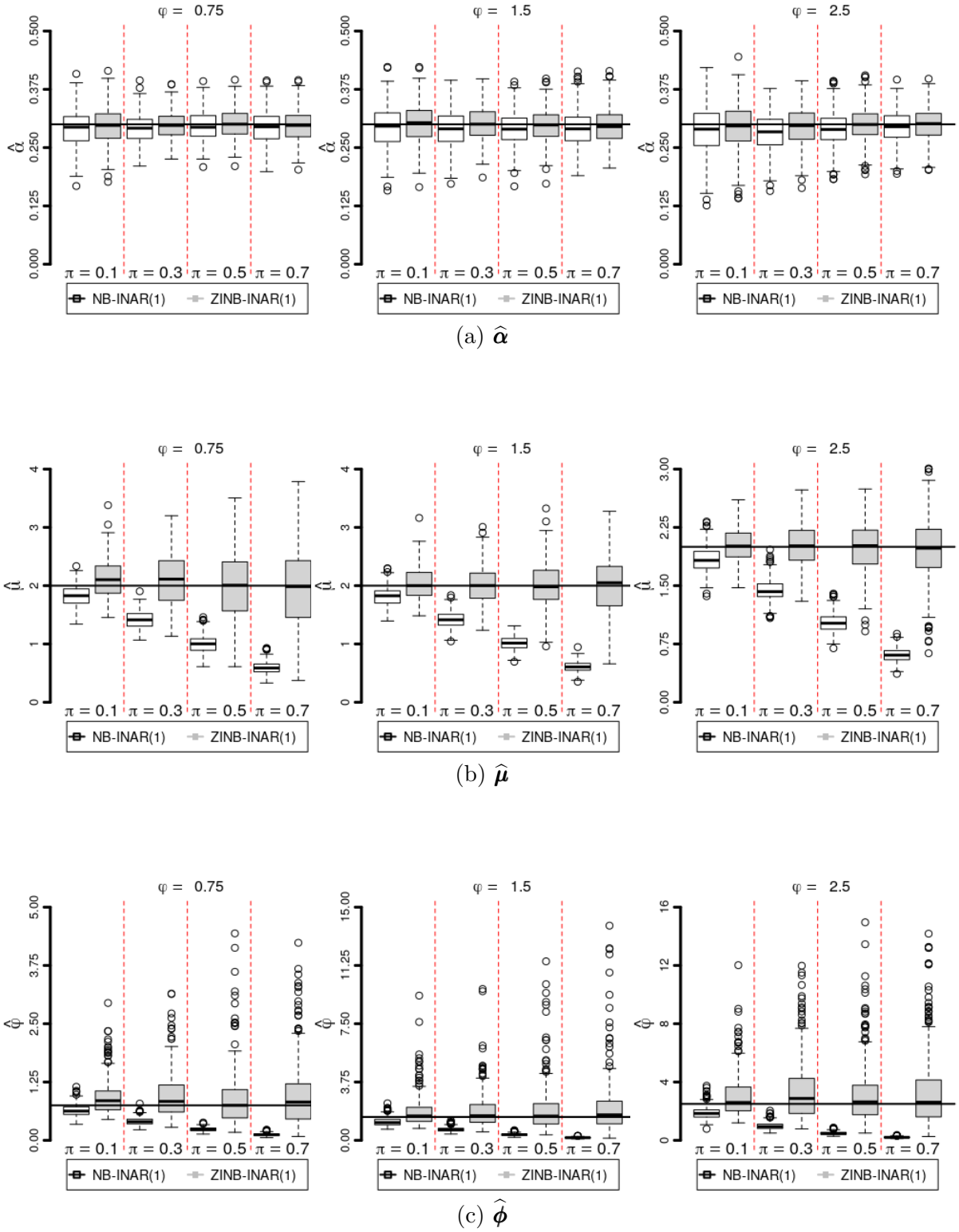


Figura 8 – Box-plot das estimativas dos parâmetros para os processos NB-INAR(1) e ZINB-INAR(1) considerando $\mu = 2$ e $n = 300$. Linha horizontal contínua preta: verdadeiro valor do parâmetro.

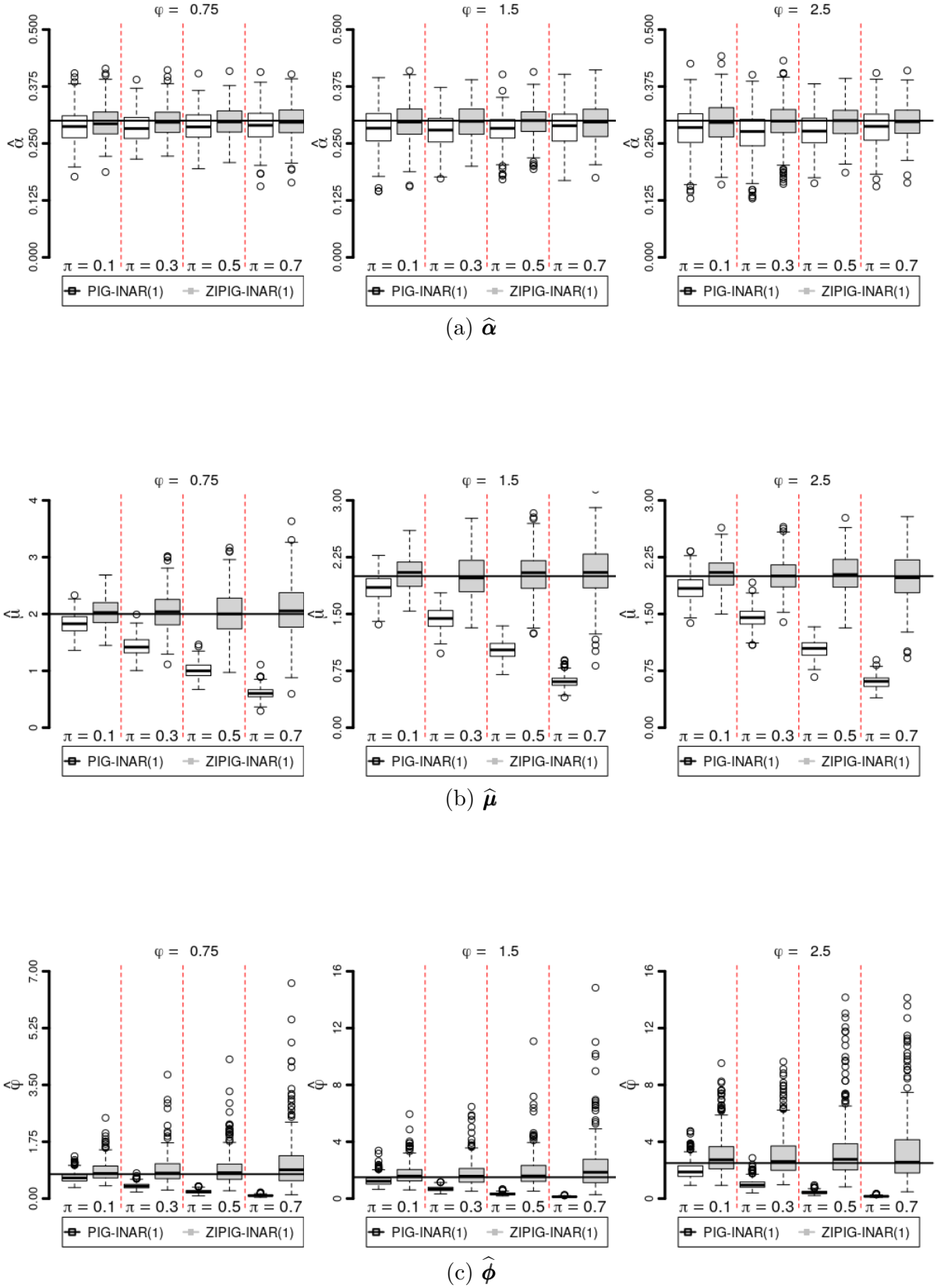


Figura 9 – Box-plot das estimativas dos parâmetros para os processos PIG-INAR(1) e ZIPIG-INAR(1) considerando $\mu = 2$ e $n = 300$. Linha horizontal contínua preta: verdadeiro valor do parâmetro.

APÊNDICE B – DETALHES DO ALGORITMO EM DO PROCESSO ZIPIG-INAR

Neste apêndice, mostramos como obter os valores esperados $\widehat{b_t z_t}^{(k)}$ e $\widehat{b_t/z_t}^{(k)}$ dados em (3.37) e (3.38) no passo E do algoritmo EM para o processo ZIPIG-INAR(1). Inicialmente, apresentamos alguns resultados envolvendo a distribuição Inversa Gaussiana Generalizada, os quais serão úteis para calcular os valores esperados em questão.

Distribuição Inversa Gaussiana Generalizada:

Seja Z uma v.a. contínua com distribuição Inversa Gaussiana Generalizada de parâmetros λ , μ e ϕ , denotada por $Z \sim \text{GIG}(\lambda, \mu, \phi)$, então sua fdp é dada por:

$$g(z; \lambda, \mu, \phi) = \mu^\lambda \frac{z^{\lambda-1}}{2K_\lambda(\mu\phi)} \exp\left\{-\frac{\phi(z-\mu)^2}{2z}\right\}.$$

Em particular, se $\lambda = -1/2$, então a v.a. Z segue distribuição Inversa Gaussiana, denotada por $\text{IG}(\mu, \phi)$. Os momentos em torno da origem da distribuição GIG são:

$$E(Z^r) = \mu^r \frac{K_{\lambda+r}(\mu\phi)}{K_\lambda(\mu\phi)}, \quad (\text{B.1})$$

para $r = \dots, -1, 0, 1, \dots$, em que $K_\lambda(t)$ a função de Bessel de terceira ordem definida no Capítulo 2. Segue algumas fórmulas envolvendo a função Bessel de terceira ordem (ABRAMOWITZ; STEGUN, 1964):

$$K_{-n}(x) = K_n(x), \quad K_{n+1}(x) = \frac{2n}{x} K_n(x) + K_{n-1}(x) \quad (\text{B.2})$$

Pode-se mostrar, utilizando as fórmulas (B.1) e (B.2) que, se $Z \sim \text{IG}(\mu, \phi)$, então:

$$E(Z) = \mu \quad \text{e} \quad E(Z^{-1}) = \frac{1}{\mu} \left(1 + \frac{1}{\mu\phi}\right) \quad (\text{B.3})$$

No contexto Bayesiano, a distribuição $\text{GIG}(\lambda, \mu, \phi)$ é conjugada da distribuição Poisson. Isto é, se $\theta \sim \text{GIG}(\lambda, \mu, \phi)$, então $\theta|V=v, t \sim \text{GIG}\left(\lambda+v, \mu\sqrt{\frac{\phi}{2t+\phi}}, 2t+\phi\right)$, sendo $V|\theta, t \sim \text{Poisson}(t\theta)$. Note que V segue uma distribuição de mistura Poisson. Para mais detalhes da distribuição GIG, ver Jorgensen (2012).

O lema a seguir, mostra a forma da esperança da posteriori da distribuição de mistura Poisson, que tem como caso particular a distribuição FIG.

Lema B.1. (SAPATINAS, 1995) A esperança a posteriori $E(\theta^r|v,t)$, em que $V|\theta, t \sim \text{Poisson}(t\theta)$ e $\theta \sim G(\theta)$ é dada por:

$$E(\theta^r|v,t) = \frac{(v+r)!P(V=v+r|t)}{\mu^r v!P(V=v|t)}, \quad r = 0, 1, \dots, \quad (\text{B.4})$$

Note que, na fórmula (B.4), o valor de r se restringe apenas para o conjunto dos valores inteiros não negativos. Em particular, para $r = -1$ e $v > 0$, temos

$$\begin{aligned} E(\theta^{-1}|v,t) &= \int_0^\infty \theta^{-1} g(\theta|v,t) d\theta \\ &= \int_0^\infty \theta^{-1} \frac{\mathbb{P}(V=v|\theta,t)g(\theta|t)}{\mathbb{P}(V=v|t)} d\theta \\ &= \frac{1}{P(V=v|t)} \int_0^\infty \theta^{-1} \frac{e^{-t\theta}(t\theta)^v}{v!} g(\theta|t) d\theta \\ &= \frac{t(v-1)!}{v!P(V=v|t)} \int_0^\infty \frac{e^{-\theta t}(\theta t)^{v-1}}{(v-1)!} g(\theta|t) d\theta \\ &= \frac{t}{v\mathbb{P}(V=v|t)} \int_0^\infty \mathbb{P}(V=v-1|\theta,t) g(\theta|t) d\theta \\ &= \frac{t\mathbb{P}(V=v-1|t)}{v\mathbb{P}(V=v|t)} \end{aligned} \quad (\text{B.5})$$

Algoritmo EM: ZIPIG-INAR(1):

com o intuito de obter $\widehat{b_t z_t}^{(k)}$ e $\widehat{b_t/z_t}^{(k)}$ no passo E, apresentamos o seguinte

Lema:

Lema B.2. Os valores esperados $E\left(B_t Z_t | V_t = v, \hat{\lambda}^{(k)}\right)$ e $E\left(B_t Z_t^{-1} | V_t = v, \hat{\lambda}^{(k)}\right)$, com $v \in \mathbb{N}$, podem ser expressos por:

$$E\left(B_t Z_t | V_t = v, \hat{\lambda}^{(k)}\right) = \frac{(1 - \hat{\pi}^{(k)}) (v+1) h\left(v+1 | \hat{\lambda}^{(k)}\right)}{\hat{\mu}^{(k)} \left[\hat{\pi}^{(k)} \mathbb{I}_{\{v=0\}} + (1 - \hat{\pi}^{(k)}) h\left(v | \hat{\lambda}^{(k)}\right) \right]} \quad (\text{B.6})$$

e

$$E\left(B_t Z_t^{-1} | V_t = v, \hat{\lambda}^{(k)}\right) = \begin{cases} \frac{\hat{\mu}^{(k)} h\left(v-1 | \hat{\lambda}^{(k)}\right)}{v h\left(v | \hat{\lambda}^{(k)}\right)}, & \text{se } v > 0 \\ \frac{\left(\sqrt{\hat{\phi}^{(k)}(\hat{\phi}^{(k)} + 2\hat{\mu}^{(k)})} + 1\right) (1 - \hat{\pi}^{(k)}) h\left(0 | \hat{\lambda}^{(k)}\right)}{\hat{\phi}^{(k)} \left[\hat{\pi}^{(k)} + (1 - \hat{\pi}^{(k)}) h\left(0 | \hat{\lambda}^{(k)}\right) \right]}, & \text{se } v = 0. \end{cases} \quad (\text{B.7})$$

Prova. Para $r \in \mathbb{Z}$, temos que:

$$\begin{aligned}
E\left(B_t Z_t^r | V_t = v, \hat{\lambda}^{(k)}\right) &= E\left[E\left(B_t Z_t^r | W_t, V_t = v, \hat{\lambda}^{(k)}\right)\right] \\
&= E\left[B_t E\left(Z_t^r | W_t, V_t = v, \hat{\lambda}^{(k)}\right) | V_t = v, \hat{\lambda}^{(k)}\right] \\
&= \mathbb{P}\left(W_t = 0 | V_t = v, \hat{\lambda}^{(k)}\right) E\left(Z_t^r | W_t = 0, V_t = v, \hat{\lambda}^{(k)}\right) \\
&= \frac{\mathbb{P}\left(W_t = 0 | \hat{\lambda}^{(k)}\right) \times \mathbb{P}\left(V_t = v | W_t = 0, \hat{\lambda}^{(k)}\right)}{\sum_{w_t=0}^1 \mathbb{P}\left(W_t = w_t | \hat{\lambda}^{(k)}\right) \times \mathbb{P}\left(V_t = v | W_t = w_t, \hat{\lambda}^{(k)}\right)} \times E\left(Z_t^r | W_t = 0, V_t = v, \hat{\lambda}^{(k)}\right)
\end{aligned} \tag{B.8}$$

$$= \frac{(1 - \hat{\pi}^{(k)}) h(v | \hat{\lambda}^{(k)})}{\hat{\pi}^{(k)} \mathbb{I}_{\{v=0\}} + (1 - \hat{\pi}^{(k)}) h(v | \hat{\lambda}^{(k)})} \times E\left(Z_t^r | W_t = 0, V_t = v, \hat{\lambda}^{(k)}\right), \tag{B.9}$$

em que (B.8) é obtido pelo Teorema de Bayes.

Note por (B.4) que

$$E\left(Z_t | W_t = 0, V_t = v, \hat{\lambda}^{(k)}\right) = \frac{v+1}{\hat{\mu}^{(k)}} \frac{h(v+1 | \hat{\lambda}^{(k)})}{h(v | \hat{\lambda}^{(k)})}. \tag{B.10}$$

Assim, obtemos (B.6) pela substituição na esperança de (B.9) por (B.10) considerando $r = 1$.

Agora, para obter $E\left(B_t Z_t^{-1} | V_t = v, \hat{\lambda}^{(k)}\right)$, consideramos os seguintes casos:

Caso 1: $v > 0$

Para $v > 0$, temos que

$$E\left(B_t Z_t^{-1} | V_t = v, \hat{\lambda}^{(k)}\right) = 1 \times E\left(Z_t^{-1} | W_t = 0, V_t = v, \hat{\lambda}^{(k)}\right) \tag{B.11}$$

$$= \frac{\hat{\mu}^{(k)} h(v-1 | \hat{\lambda}^{(k)})}{v h(v | \hat{\lambda}^{(k)})}, \tag{B.12}$$

em que (B.11) é obtido por (B.9), substituindo a função indicadora por zero, e (B.12) é obtido imediatamente de (B.5).

Caso 2: $v = 0$.

Neste caso, temos $V_t = 0$. Note que, no contexto bayesiano, a distribuição de $Z_t | W_t = 0$ é a priori conjugada da distribuição Poisson. Temos então que,

$$Z_t | W_t = 0, \hat{\lambda}^{(k)} \sim IG(1, \hat{\phi}^{(k)})$$

e

$$Z_t|V_t=0, W_t=0, \hat{\lambda}^{(k)} \sim IG\left(\sqrt{\frac{\hat{\phi}^{(k)}}{2\hat{\mu}^{(k)} + \hat{\phi}^{(k)}}}, 2\hat{\mu}^{(k)} + \hat{\phi}^{(k)}\right)$$

Portanto, por (B.3), temos:

$$E\left(Z_t^{-1}|V_t=0, W_t=0, \hat{\lambda}^{(k)}\right) = \frac{\sqrt{\hat{\phi}^{(k)}(\hat{\phi}^{(k)} + 2\hat{\mu}^{(k)})} + 1}{\hat{\phi}^{(k)}}$$

Logo, de (B.9) para $r = -1$, obtemos

$$E\left(B_t Z_t^{-1}|V_t=0, W_t=0, \hat{\lambda}^{(k)}\right) = \frac{\left(\sqrt{\hat{\phi}^{(k)}(\hat{\phi}^{(k)} + 2\hat{\mu}^{(k)})} + 1\right) \left(1 - \hat{\pi}^{(k)}\right) h\left(0|\hat{\lambda}^{(k)}\right)}{\hat{\phi}^{(k)} \left[\hat{\pi}^{(k)} + \left(1 - \hat{\pi}^{(k)}\right) h\left(0|\hat{\lambda}^{(k)}\right)\right]} \quad (\text{B.13})$$

Utilizando os casos 1 e 2, temos por (B.12) e (B.13) que a expressão (B.7) é verdadeira.

O seguinte resultado também é útil para a demonstração dos valores esperados em questão:

$$\begin{aligned} E\left[E\left(B_t Z_t^r|S_t, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\lambda}^{(k)}\right)\right] &= \\ &= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} E\left(B_t Z_t^r|S_t = s_t, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\lambda}^{(k)}\right) \times \mathbb{P}\left(S_t = s_t|Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\lambda}^{(k)}\right) \\ &= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} E\left(B_t Z_t^r|V_t = y_t - s_t, \hat{\lambda}^{(k)}\right) \times \mathbb{P}\left(S_t = s_t|Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\lambda}^{(k)}\right) \end{aligned} \quad (\text{B.14})$$

em que (B.14) é obtido utilizando o seguinte resultado:

$$E\left(B_t Z_t^r|S_t = s_t, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\lambda}^{(k)}\right) = E\left(B_t Z_t^r|V_t = y_t - s_t, \hat{\lambda}^{(k)}\right), \quad (\text{B.15})$$

do qual é válido uma vez que (B_t, Z_t) depende apenas de V_t , que condicionado a $Y_t = y_t$ e $S_t = s_t$, é determinado por $V_t = y_t - s_t$.

Assim, considerando $r = 1$ e $r = -1$, e utilizando o Lema (B.2) no resultado (B.14), obtemos:

$$\begin{aligned}
\widehat{b_t z_t}^{(k)} &= E \left[E \left(B_t Z_t | S_t, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\lambda}^{(k)} \right) \right] \\
&= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} E \left(B_t Z_t | V_t = y_t - s_t, \hat{\lambda}^{(k)} \right) \times \mathbb{P} \left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\lambda}^{(k)} \right) \\
&= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \frac{(1 - \hat{\pi}^{(k)}) (y_t - s_t + 1) h \left(y_t - s_t + 1 | \hat{\lambda}^{(k)} \right) \mathbb{P} \left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\theta}^{(k)} \right)}{\hat{\mu}^{(k)} \left[\hat{\pi}^{(k)} \mathbb{I}_{\{y_t - s_t = 0\}} + (1 - \hat{\pi}^{(k)}) h \left(y_t - s_t | \hat{\lambda}^{(k)} \right) \right]}
\end{aligned}$$

e

$$\begin{aligned}
\widehat{b_t / z_t}^{(k)} &= E \left[E \left(B_t Z_t^{-1} | S_t, Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\lambda}^{(k)} \right) \right] \\
&= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} E \left(B_t Z_t^{-1} | V_t = y_t - s_t, \hat{\lambda}^{(k)} \right) \times \mathbb{P} \left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\lambda}^{(k)} \right) \\
&= \sum_{s_t=0}^{\min\{y_{t-1}, y_t\}} \left\{ \frac{\hat{\mu}^{(k)} h \left(y_t - s_t - 1 | \hat{\lambda}^{(k)} \right) \times \mathbb{P} \left(S_t = s_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\theta}^{(k)} \right)}{(y_t - s_t) h \left(y_t - s_t | \hat{\lambda}^{(k)} \right)} \mathbb{I}_{\{y_t - s_t > 0\}} \right. \\
&\quad \left. + \frac{\left(\sqrt{\hat{\phi}^{(k)} (\hat{\phi}^{(k)} + 2\hat{\mu}^{(k)})} + 1 \right) (1 - \hat{\pi}^{(k)}) h \left(0 | \hat{\lambda}^{(k)} \right) \times \mathbb{P} \left(S_t = y_t | Y_t = y_t, Y_{t-1} = y_{t-1}, \hat{\theta}^{(k)} \right)}{\hat{\phi}^{(k)} \left[\hat{\pi}^{(k)} + (1 - \hat{\pi}^{(k)}) h \left(0 | \hat{\lambda}^{(k)} \right) \right]} \mathbb{I}_{\{y_t - s_t = 0\}} \right\},
\end{aligned}$$

que coincide com as expressões de $\widehat{b_t z_t}^{(k)}$ e $\widehat{b_t / z_t}^{(k)}$ fornecidas em (3.37) e (3.38), respectivamente.

Algoritmo EM: ZIPIG-INAR(p):

Para encontrar as expressões de $\widehat{b_t z_t}^{(k)}$ e $\widehat{b_t / z_t}^{(k)}$ considerando a ordem p do processo ZIPIG-INAR, apresentamos o seguinte resultado:

$$E \left(B_t Z_t^r | \mathbf{S}_t = \mathbf{s}_t, Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\lambda}^{(k)} \right) = E \left(B_t Z_t^r | V_t = y_t - \sum_{i=1}^p s_{t,i}, \hat{\lambda}^{(k)} \right) \quad (\text{B.16})$$

O resultado (B.16) é uma generalização de (B.15) para ordem p e é válido por (B_t, Z_t) depender apenas de V_t que, dado $Y_t = y_t$ e $\mathbf{S}_t = \mathbf{s}_t$, é obtido por $V_t = y_t - \sum_{i=1}^p s_{t,i}$. Assim, para $r = 1$ e $r = -1$, obtemos:

$$\begin{aligned}
\widehat{b_t z_t}^{(k)} &= E \left[E \left(B_t Z_t | \mathbf{S}_t, Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\lambda}^{(k)} \right) \right] \\
&= \sum_{s=0}^{\min\{\sum_{i=1}^p y_{t-i}, y_t\}} E \left(B_t Z_t | V_t = y_t - s, \hat{\lambda}^{(k)} \right) \times \mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\lambda}^{(k)} \right) \\
&= \sum_{s=0}^{\min\{\sum_{i=1}^p y_{t-i}, y_t\}} \frac{(1 - \hat{\pi}^{(k)}) (y_t - s + 1) h \left(y_t - s + 1 | \hat{\lambda}^{(k)} \right) \mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\theta}^{(k)} \right)}{\hat{\mu}^{(k)} \left[\hat{\pi}^{(k)} \mathbb{I}_{\{y_t = s\}} + (1 - \hat{\pi}^{(k)}) h \left(y_t - s | \hat{\lambda}^{(k)} \right) \right]}
\end{aligned}$$

e

$$\begin{aligned}
\widehat{b_t/z_t}^{(k)} &= E \left[E \left(B_t Z_t^{-1} | \mathbf{S}_t, Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\lambda}}^{(k)} \right) \right] \\
&= \sum_{s=0}^{\min\{\sum_{i=1}^p y_{t-i}, y_t\}} E \left(B_t Z_t^{-1} | V_t = y_t - s, \hat{\boldsymbol{\lambda}}^{(k)} \right) \times \mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\lambda}}^{(k)} \right) \\
&= \sum_{s=0}^{\min\{\sum_{i=1}^p y_{t-i}, y_t\}} \left\{ \frac{\hat{\mu}^{(k)} h \left(y_t - s - 1 | \hat{\boldsymbol{\lambda}}^{(k)} \right) \times \mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)} \right)}{(y_t - s) h \left(y_t - s | \hat{\boldsymbol{\lambda}}^{(k)} \right)} \mathbb{I}_{\{y_t > s\}} \right. \\
&\quad \left. + \frac{\left(\sqrt{\hat{\phi}^{(k)}(\hat{\phi}^{(k)} + 2\hat{\mu}^{(k)})} + 1 \right) (1 - \hat{\pi}^{(k)}) h \left(0 | \hat{\boldsymbol{\lambda}}^{(k)} \right) \times \mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)} \right)}{\hat{\phi}^{(k)} \left[\hat{\pi}^{(k)} + (1 - \hat{\pi}^{(k)}) h \left(0 | \hat{\boldsymbol{\lambda}}^{(k)} \right) \right]} \mathbb{I}_{\{y_t = s\}} \right\},,
\end{aligned}$$

em que, pelo Teorema de Bayes, a probabilidade do numerador é obtida como:

$$\mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | Y_t = y_t, \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)} \right) = \frac{\mathbb{P} \left(\sum_{i=1}^p S_{t,i} = s | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)} \right) \times \mathbb{P} \left(V_t = y_t - s | \hat{\boldsymbol{\theta}}^{(k)} \right)}{\mathbb{P} \left(Y_t = y_t | \mathbf{Y}_{t(-p)} = \mathbf{y}_{t(-p)}, \hat{\boldsymbol{\theta}}^{(k)} \right)},$$

que é uma função das probabilidades dadas em (4.12) e (4.24).

APÊNDICE C – DEMONSTRAÇÃO DO TEOREMA 4.1

Prova do item i):

Para demonstrar o item i) do Teorema 4.1, precisamos provar as seguintes afirmações:

(1) Se $\alpha_i \in [0, 1)$, para $i = 1, \dots, p$, e $\sum_{i=1}^p \alpha_i < 1$, então $\beta_i \in [0, 1)$, para $i = 1, \dots, p$.

(2) Se $\beta_i \in [0, 1)$, para $i = 1, \dots, p$, então $\alpha_i \in [0, 1)$, para $i = 1, \dots, p$, e $\sum_{i=1}^p \alpha_i < 1$.

Prova da afirmação (1): Dado $\alpha_i \in [0, 1)$, para $i = 1, \dots, p$, e $\sum_{i=1}^p \alpha_i < 1$, temos que $\alpha_i < 1 - \sum_{j \neq i}^p \alpha_j$. Logo, $\beta_i = \frac{\alpha_i}{1 - \sum_{j \neq i}^p \alpha_j} < 1$ e ainda $\beta_i \geq 0$, uma vez que $0 \leq \alpha_i < 1 - \sum_{j \neq i}^p \alpha_j$.

Prova da afirmação (2): Se $\beta_i \in [0, 1)$, para $i = 1, \dots, p$, então:

$$\beta_i = \frac{\alpha_i}{1 - \sum_{j \neq i}^p \alpha_j} \Rightarrow 1 - \sum_{j \neq i}^p \alpha_j > \alpha_i \Rightarrow \sum_{j=1}^p \alpha_j < 1.$$

Agora, note que:

$$0 \leq \beta_i < 1 \Rightarrow 0 \leq \frac{\alpha_i}{1 - \sum_{j \neq i}^p \alpha_j} < 1 \Rightarrow 0 \leq 1 - \frac{\alpha_i}{1 - \sum_{j \neq i}^p \alpha_j} < 1 \Rightarrow 0 \leq \frac{1 - \sum_{j=1}^p \alpha_j}{1 - \sum_{j=1}^p \alpha_j + \alpha_i} < 1. \quad (\text{C.1})$$

Uma vez que $1 - \sum_{j=1}^p \alpha_j > 0$, temos que (C.1) implica necessariamente que $\alpha_i \geq 0$, para $i = 1, \dots, p$. □

Prova do item i):

Note que a equação (4.25) pode ser reescrita na forma matricial $A(\boldsymbol{\beta})\boldsymbol{\alpha} = \boldsymbol{\beta}$, em que:

$$A(\boldsymbol{\beta}) = \begin{pmatrix} 1 & \beta_1 & \beta_1 & \dots & \beta_1 \\ \beta_2 & 1 & \beta_2 & \dots & \beta_2 \\ \beta_3 & \beta_3 & 1 & \dots & \beta_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_p & \beta_p & \beta_p & \dots & 1 \end{pmatrix}, \quad (\text{C.2})$$

e $\beta = (\beta_1, \dots, \beta_p)^\top$, com $\beta_i \in [0, 1)$, para $i = 1, \dots, p$. Assim, para demonstrar o Teorema 4.1, desenvolvemos a expressão do cofator e do determinante da matriz $A(\beta)$, dos quais serão utilizados para encontrar a expressão de α_i . Ao longo da demonstração, adotaremos também a notação $A(\beta_1, \dots, \beta_p)$ para representar a matriz $A(\beta)$.

O lema C.1 é útil para desenvolver a expressão do cofator de $A(\beta)$. Este lema afirma que o determinante da matriz $A(\beta)$ é invariante por permutação do vetor β

Lema C.1. *Seja $\beta = (\beta_1, \dots, \beta_p)^\top$, com $\beta_i \in [0, 1)$, para $i = 1, \dots, p$, e β^* uma permutação do vetor β , isto é, $\beta^* = (\beta_{i_1}, \dots, \beta_{i_p})$, $1 \leq i_1 \neq i_2 \neq \dots \neq i_p \leq p$. Então, $\det(A(\beta^*)) = \det(A(\beta))$, em que $A(\cdot)$ é definido em (C.2).*

Prova. *A ideia para demonstrar o lema é provar que $A(\beta^*)$ pode ser obtida permutando linhas e colunas da matriz $A(\beta)$. Assim, seja $l, k \in \{1, \dots, p\}$, com $l \neq k$. Obtenha uma matriz $p \times p$ utilizando os seguintes passos:*

1. *Permute a coluna l pela coluna k da matriz $A(\beta)$.*
2. *Permute a linha l pela linha k da matriz resultante do passo anterior.*

Note que a matriz resultante dos passos 1 e 2 é a matriz $A(\tilde{\beta})$, em que $\tilde{\beta}$ é o vetor β com os elementos das posições l e k trocados, isto é, $\tilde{\beta} = (\beta_1, \dots, \beta_{k-1}, \beta_l, \beta_{k+1}, \dots, \beta_{l-1}, \beta_k, \beta_{l+1}, \dots, \beta_p)$. Utilizando a propriedade de troca de linhas e colunas do determinante, é fácil ver que:

$$\det(A(\beta)) = \det(A(\tilde{\beta})). \quad (\text{C.3})$$

Assim, realizando sucessivas trocas de elementos de β que resultam no vetor β^ , temos por (C.3) que o determinante da matriz não será alterado em cada permutação. Logo, $\det(A(\beta^*)) = \det(A(\beta))$. $\square$*

Proposição C.1. *Seja $\beta = (\beta_1, \dots, \beta_p)^\top$. Seja β_{-i} o vetor β sem o i -ésimo elemento e $\beta_{-i,-j}$ o vetor β sem o i -ésimo e j -ésimo elemento. Então, o cofator da matriz $A(\cdot)$ definida em (C.2) é dado por:*

$$\text{cof}_{i,j}(A(\beta)) = \begin{cases} \det(A(\beta_{-i})), & \text{se } i = j \\ -\beta_j \det(A(\beta_{-i,-j}, 1)), & \text{se } i \neq j, \end{cases} \quad (\text{C.4})$$

em que $A(\beta, 1) = A(\beta_1, \dots, \beta_p, 1)$ e a seguinte relação é satisfeita:

$$\det(A(\beta, 1)) = \prod_{i=1}^p (1 - \beta_i). \quad (\text{C.5})$$

Prova. Inicialmente, mostramos a validade de (C.4), separando a demonstração em três casos:

Caso 1: Seja $i = j$. Neste caso, é fácil ver que a matriz resultante da eliminação da i -ésima linha e i -ésima coluna $A(\beta)$ é a matriz $A(\beta_{-i})$. Portanto,

$$\text{cof}_{i,i}(A(\beta)) = (-1)^{i+i} \det(A(\beta_{-i})) = \det(A(\beta_{-i}))$$

Caso 2: Seja $i < j$. Neste caso, temos:

$$\text{cof}_{i,j}(A(\beta)) = (-1)^{i+j} \begin{vmatrix} & i-1 & i & i+1 & & j-1 & & j \\ & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \dots & 1 & \beta_{i-1} & \beta_{i-1} & \dots & \beta_{i-1} & \beta_{i-1} & \beta_{i-1} & \dots & i-1 \\ \hline & \beta_i & 1 & \beta_i & \dots & \beta_i & \beta_i & \beta_i & \dots \\ \dots & \beta_{i+1} & \beta_{i+1} & 1 & \dots & \beta_{i+1} & \beta_{i+1} & \beta_{i+1} & \dots & i \\ & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & \\ \dots & \beta_{j-1} & \beta_{j-1} & \beta_{j-1} & \dots & 1 & \beta_{j-1} & \beta_{j-1} & \dots & j-2 \\ \dots & \beta_j & \beta_j & \beta_j & \dots & \beta_j & 1 & \beta_j & \dots & j-1 \\ \dots & \beta_{j+1} & \beta_{j+1} & \beta_{j+1} & \dots & \beta_{j+1} & \beta_{j+1} & 1 & \dots & j \\ & \vdots & \vdots & \vdots & & \vdots & & \vdots & \end{vmatrix} \quad (\text{C.6})$$

Permutando na matriz em (C.6) as linhas $j-1$ e $j-2$, em seguida, as linhas $j-2$ e $j-3$, e assim sucessivamente, até permutar as linhas $i+1$ e i , obtemos:

$$\text{cof}_{i,j}(A(\beta)) = (-1)^{i+j} (-1)^{(j-1)-i} \begin{vmatrix} & i-1 & i & i+1 & & j-1 & & j \\ & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \dots & 1 & \beta_{i-1} & \beta_{i-1} & \dots & \beta_{i-1} & \beta_{i-1} & \beta_{i-1} & \dots & i-1 \\ \hline & \beta_i & 1 & \beta_i & \dots & \beta_i & \beta_i & \beta_i & \dots \\ \dots & \beta_j & \beta_j & \beta_j & \dots & \beta_j & 1 & \beta_j & \dots & i \\ & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & \\ \dots & \beta_{j-2} & \beta_{j-2} & \beta_{j-2} & \dots & \beta_{j-2} & \beta_{j-2} & \beta_{j-2} & \dots & j-2 \\ \dots & \beta_{j-1} & \beta_{j-1} & \beta_{j-1} & \dots & 1 & \beta_{j-1} & \beta_{j-1} & \dots & j-1 \\ \dots & \beta_{j+1} & \beta_{j+1} & \beta_{j+1} & \dots & \beta_{j+1} & \beta_{j+1} & 1 & \dots & j \\ & \vdots & \vdots & \vdots & & \vdots & & \vdots & \end{vmatrix} \quad (\text{C.7})$$

Note que todos os elementos da i -ésima linha da matriz (C.7) são iguais a β_j . Note ainda que, substituindo estes β_j 's por 1 a matriz em (C.7) se transforma na matriz $A(\beta_1, \dots, \beta_{i-1}, 1, \beta_{i+1}, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_p)$. Portanto, utilizando propriedades do determinante e o resultado da proposição C.1, obtemos:

$$\begin{aligned} \text{cof}_{i,j}(A(\beta)) &= -\beta_j \det(A(\beta_1, \dots, \beta_{i-1}, 1, \beta_{i+1}, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_p)) \\ &= -\beta_j \det(A(\beta_{-i,-j}, 1)), \end{aligned}$$

Caso 3: Seja $i > j$. Neste caso, temos:

$$\text{cof}_{i,j}(A(\boldsymbol{\beta})) = (-1)^{i+j} \begin{vmatrix} & j-1 & & j & & i-2 & i-1 & i \\ & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & 1 & \beta_{j-1} & \beta_{j-1} & \dots & \beta_{j-1} & \beta_{j-1} & \beta_{j-1} & \dots & j-1 \\ \dots & \beta_j & 1 & \beta_j & \dots & \beta_j & \beta_j & \beta_j & \dots & j \\ \dots & \beta_{j+1} & \beta_{j+1} & 1 & \dots & \beta_{j+1} & \beta_{j+1} & \beta_{j+1} & \dots & j+1 \\ & \vdots & & \vdots & \ddots & \vdots & \vdots & \vdots & & \\ \dots & \beta_{i-1} & \beta_{i-1} & \beta_{i-1} & \dots & 1 & \beta_{i-1} & \beta_{i-1} & \dots & i-1 \\ \hline \dots & \beta_i & \beta_i & \beta_i & \dots & \beta_i & 1 & \beta_i & \dots & \\ \dots & \beta_{i+1} & \beta_{i+1} & \beta_{i+1} & \dots & \beta_{i+1} & \beta_{i+1} & 1 & \dots & i \\ & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & & \end{vmatrix} \quad (\text{C.8})$$

Permutando na matriz em (C.8) as linhas j e $j+1$, em seguida, as linhas $j+1$ e $j+2$, e assim sucessivamente, até permutar as linhas $i-2$ e $i-1$, obtemos:

$$\text{cof}_{i,j}(A(\boldsymbol{\beta})) = (-1)^{i+j}(-1)^{(i-1)-j} \begin{vmatrix} & j-1 & & j & & i-2 & i-1 & i \\ & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & 1 & \beta_{j-1} & \beta_{j-1} & \dots & \beta_{j-1} & \beta_{j-1} & \beta_{j-1} & \dots & j-1 \\ \dots & \beta_{j+1} & \beta_{j+1} & 1 & \dots & \beta_{j+1} & \beta_{j+1} & \beta_{j+1} & \dots & j \\ \dots & \beta_{j+2} & \beta_{j+2} & \beta_{j+2} & \dots & \beta_{j+2} & \beta_{j+2} & \beta_{j+2} & \dots & j+1 \\ & \vdots & & \vdots & \ddots & \vdots & \vdots & \vdots & & \\ \dots & \beta_j & 1 & \beta_j & \dots & \beta_j & \beta_j & \beta_j & \dots & i-1 \\ \hline \dots & \beta_i & \beta_i & \beta_i & \dots & \beta_i & 1 & \beta_i & \dots & \\ \dots & \beta_{i+1} & \beta_{i+1} & \beta_{i+1} & \dots & \beta_{i+1} & \beta_{i+1} & 1 & \dots & i \\ & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & & \end{vmatrix}$$

em que todos os elementos da linha $i-1$ são iguais a β_j . Assim, analogamente a demonstração do caso 2, obtemos:

$$\begin{aligned} \text{cof}_{i,j}(A(\boldsymbol{\beta})) &= -\beta_j \det(A(\beta_1, \dots, \beta_{j-1}, 1, \beta_{j+1}, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_p)) \\ &= -\beta_j \det(A(\boldsymbol{\beta}_{-i,-j}, 1)), \end{aligned}$$

Agora, demonstraremos a equação (C.5).

Seja $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}, 1)$ e a_{ij} , $i, j \in \{1, \dots, p+1\}$ os elementos da matriz $A(\tilde{\boldsymbol{\beta}})$. Então, pelo teorema de Laplace e por (C.4), temos:

$$\begin{aligned}
\det(A(\tilde{\beta})) &= \sum_{j=1}^{p+1} a_{1,j} \text{cof}_{1,j}(A(\tilde{\beta})) \\
&= 1 \times \text{cof}_{1,1}(A(\tilde{\beta})) - \sum_{j=2}^p \beta_1 \text{cof}_{1,j}(A(\tilde{\beta})) - \beta_1 \text{cof}_{1,p+1}(A(\tilde{\beta})) \\
&= \det(A(\tilde{\beta}_{-1})) - \beta_1 \sum_{j=2}^p \tilde{\beta}_j \det(A(\tilde{\beta}_{-1,-j}, 1)) - \beta_1 \tilde{\beta}_{p+1} \det(A(\tilde{\beta}_{-1,-(p+1)}, 1)) \\
&= \det(A(\beta_2, \dots, \beta_p, 1)) - \beta_1 \sum_{j=2}^p \beta_j \det(A(\beta_2, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_p, 1, 1)) \\
&\quad - \beta_1 \times 1 \times \det(A(\beta_2, \dots, \beta_p, 1)) \tag{C.9}
\end{aligned}$$

$$= \det(A(\beta_2, \dots, \beta_p, 1))(1 - \beta_1), \tag{C.10}$$

em que utilizamos o fato que $\tilde{\beta}_{p+1} = 1$ e $\tilde{\beta}_j = \beta_j$, com $j = 1, \dots, p$, para obter (C.9) e a expressão (C.10) foi obtida por consequência de que $\det(A(\beta_2, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_p, 1, 1)) = 0$, uma vez que a matriz $A(\beta_2, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_p, 1, 1)$ tem duas linhas com todos os elementos iguais a 1. Assim, de (C.10), podemos obter $\det(A(\tilde{\beta}))$ de forma recursiva da seguinte maneira:

$$\begin{aligned}
\det(A(\tilde{\beta})) &= \det(A(\beta_1, \dots, \beta_p, 1)) \\
&= \det(A(\beta_2, \dots, \beta_p, 1))(1 - \beta_1) \\
&= \det(A(\beta_3, \dots, \beta_p, 1))(1 - \beta_2)(1 - \beta_1) \\
&\quad \vdots \\
&= \det(A(\beta_p, 1)) \prod_{j=1}^{p-1} (1 - \beta_j) = \begin{vmatrix} 1 & \beta_p \\ 1 & 1 \end{vmatrix}^{p-1} \prod_{j=1}^{p-1} (1 - \beta_j) = \prod_{j=1}^p (1 - \beta_j) \quad \square
\end{aligned}$$

Proposição C.2. Seja $\beta = (\beta_1, \dots, \beta_p)$, com $\beta_i \in [0, 1)$, para $i = 1, \dots, p$. Então, o determinante da matriz $A(\beta)$ definida em (C.2) é dado por:

$$\det(A(\beta)) = 1 - \sum_{i=1}^{p-1} \sum_{j=i+1}^p \beta_i \left(\frac{\beta_j}{1 - \beta_j} \right) \prod_{l=i+1}^p (1 - \beta_l) \tag{C.11}$$

Prova. Para demonstrar a validade da expressão (C.11) utilizaremos indução. Supondo $p = 2$, então $\beta = (\beta_1, \beta_2)$ e $\det(A(\beta_1, \beta_2)) = 1 - \beta_1 \beta_2$, que coincide com (C.11). Supondo que a Proposição C.2 seja válida para $p = m - 1$ (hipótese de indução), então, para $p = m$, e denotando por $a_{i,j}$ os elementos da matriz $A(\beta)$, com $i, j \in \{1, \dots, p\}$, temos:

$$\begin{aligned}
\det(A(\boldsymbol{\beta})) &= \sum_{j=1}^m a_{1,j} \text{cof}_{1,j}(A(\boldsymbol{\beta})) \\
&= \text{cof}_{1,1}(A(\boldsymbol{\beta})) + \beta_1 \sum_{j=2}^m \text{cof}_{1,j}(A(\boldsymbol{\beta})) \\
&= \det(A(\boldsymbol{\beta}_{-1})) - \beta_1 \sum_{j=2}^m \beta_j \det(A(\boldsymbol{\beta}_{-1,-j}, 1)) \\
&= 1 - \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} (\boldsymbol{\beta}_{-1})_i \frac{(\boldsymbol{\beta}_{-1})_j}{1 - (\boldsymbol{\beta}_{-1})_j} \prod_{l=i+1}^{m-1} (1 - (\boldsymbol{\beta}_{-1})_l) - \beta_1 \sum_{j=2}^m \beta_j \frac{\prod_{l=2}^m (1 - \beta_l)}{1 - \beta_j} \\
&\quad \text{(C.12)} \\
&= 1 - \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} \beta_{i+1} \frac{\beta_{j+1}}{1 - \beta_{j+1}} \prod_{l=i+1}^{m-1} (1 - \beta_{l+1}) - \beta_1 \sum_{j=2}^m \frac{\beta_j}{1 - \beta_j} \prod_{l=2}^m (1 - \beta_l) \\
&= 1 - \sum_{i=2}^{m-1} \sum_{j=i+1}^m \beta_i \frac{\beta_j}{1 - \beta_j} \prod_{l=i+1}^m (1 - \beta_l) - \beta_1 \sum_{j=2}^m \frac{\beta_j}{1 - \beta_j} \prod_{l=2}^m (1 - \beta_l) \\
&= 1 - \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta_i \left(\frac{\beta_j}{1 - \beta_j} \right) \prod_{l=i+1}^m (1 - \beta_l),
\end{aligned}$$

em que na expressão (C.12) foi utilizado a hipótese de indução (note que $A(\boldsymbol{\beta}_{-1})$ tem $m-1$ elementos) e o resultado (C.5). $\square$

Agora, iremos demonstrar que α_i satisfaz (4.26) e que Δ é invertível. Note que, se $A(\boldsymbol{\beta})$ admite inversa, então:

$$A(\boldsymbol{\beta})\boldsymbol{\alpha} = \boldsymbol{\beta} \Rightarrow \boldsymbol{\alpha} = A^{-1}(\boldsymbol{\beta})\boldsymbol{\beta} = \frac{1}{\det(A(\boldsymbol{\beta}))} \text{cof}^T(A(\boldsymbol{\beta}))\boldsymbol{\beta}.$$

Portanto:

$$\begin{aligned}
\alpha_i &= \frac{1}{\det(A(\boldsymbol{\beta}))} \left\{ [\text{cof}^T(A(\boldsymbol{\beta}))]_{i,1}, \dots, [\text{cof}^T(A(\boldsymbol{\beta}))]_{i,p} \right\} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \\
&= \frac{1}{\det(A(\boldsymbol{\beta}))} \{ \text{cof}_{1,i}(A(\boldsymbol{\beta})), \dots, \text{cof}_{p,i}(A(\boldsymbol{\beta})) \} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \\
&= \frac{1}{\det(A(\boldsymbol{\beta}))} \sum_{k=1}^p \beta_k \text{cof}_{k,i}(A(\boldsymbol{\beta})) \\
&= \frac{1}{\det(A(\boldsymbol{\beta}))} \left[\beta_i \text{cof}_{i,i}(A(\boldsymbol{\beta})) + \sum_{k \neq i}^p \beta_k \text{cof}_{k,i}(A(\boldsymbol{\beta})) \right] \\
&= \frac{1}{\det(A(\boldsymbol{\beta}))} \left[\beta_i \text{cof}_{i,i}(A(\boldsymbol{\beta})) - \sum_{k \neq i}^p \beta_k \beta_i \det(A(\boldsymbol{\beta}_{-k,-i,1})) \right] \quad (\text{C.13})
\end{aligned}$$

$$\begin{aligned}
&= \frac{\beta_i}{\det(A(\boldsymbol{\beta}))} \left[\text{cof}_{i,i}(A(\boldsymbol{\beta})) - \sum_{k \neq i}^p \beta_k \det(A(\boldsymbol{\beta}_{-k,-i,1})) \right] \\
&= \frac{\beta_i}{\det(A(\boldsymbol{\beta}))} \left[\text{cof}_{i,i}(A(\boldsymbol{\beta})) + \sum_{k \neq i}^p \text{cof}_{i,k}(A(\boldsymbol{\beta})) \right] \\
&= \frac{\beta_i}{\det(A(\boldsymbol{\beta}))} \sum_{k=1}^p \text{cof}_{i,k}(A(\boldsymbol{\beta})) \\
&= \frac{\beta_i}{\det(A(\boldsymbol{\beta}))} \sum_{k=1}^p \text{cof}_{i,k}(A(\beta_1, \dots, \beta_{i-1}, 1, \beta_{i+1}, \dots, \beta_p)) \quad (\text{C.14})
\end{aligned}$$

$$\begin{aligned}
&= \frac{\beta_i \det(A(\beta_1, \dots, \beta_{i-1}, 1, \beta_{i+1}, \dots, \beta_p))}{\det(A(\boldsymbol{\beta}))} \\
&= \frac{\beta_i \det(A(\boldsymbol{\beta}_{-i,1}))}{\det(A(\boldsymbol{\beta}))} \quad (\text{C.15})
\end{aligned}$$

$$\begin{aligned}
&= \frac{\frac{\beta_i}{1 - \beta_i} \prod_{l=1}^p (1 - \beta_l)}{1 - \sum_{j=1}^{p-1} \sum_{k=j+1}^p \beta_j \frac{\beta_k}{1 - \beta_k} \prod_{l=k+1}^p (1 - \beta_l)}, \quad (\text{C.16})
\end{aligned}$$

em que (C.14) é obtido pela igualdade $\text{cof}_{i,k}(A(\boldsymbol{\beta})) = \text{cof}_{i,k}(A(\beta_1, \dots, \beta_{i-1}, 1, \beta_{i+1}, \dots, \beta_p))$, que é válida uma vez que as matrizes resultantes da eliminação da i -ésima linha de $A(\boldsymbol{\beta})$ e $A(\beta_1, \dots, \beta_{i-1}, 1, \beta_{i+1}, \dots, \beta_p)$ são iguais. As expressões (C.13), (C.15) e (C.16) são consequências do lema C.1 e das proposições C.1 e C.2.

Para concluir a demonstração do Teorema, mostramos que a transformação Δ

é um a um. De fato, note que:

$$\sum_{i=1}^p \alpha_i = \frac{\sum_{i=1}^p \frac{\beta_i}{1-\beta_i} \prod_{l=1}^p (1-\beta_l)}{\det(A(\boldsymbol{\beta}))} < 1.$$

Assim, como $\beta_i \in [0, 1)$, temos:

$$0 \leq \sum_{i=1}^p \frac{\beta_i}{1-\beta_i} \prod_{l=1}^p (1-\beta_l) < \det(A(\boldsymbol{\beta}))$$

Portanto, $A(\boldsymbol{\beta})$ admite inversa e consequentemente a transformação Δ é um a um. $\square$