



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE CIÊNCIAS EXATAS E DA TERRA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA
APLICADA E ESTATÍSTICA



Gráficos de Controle para o Monitoramento de Processos Autorregressivo de Valores Inteiros com Inflação ou Deflação de Zeros

Fidel Henrique Fernandes

Natal-RN

Fevereiro de 2018

Fidel Henrique Fernandes

Gráficos de Controle para o Monitoramento de Processos Autorregressivo de Valores Inteiros com Inflação ou Deflação de Zeros

Trabalho apresentado ao Programa de Pós-Graduação em Matemática Aplicada e Estatística da Universidade Federal do Rio Grande do Norte, em cumprimento com as exigências legais para obtenção do título de Mestre.

Área de Concentração: Probabilidade e Estatística.

Linha de Pesquisa: Estatística Industrial.

Orientador

Marcelo Bourguignon Pereira

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE - UFRN
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA E
ESTATÍSTICA - PPGMAE

Natal-RN

Fevereiro de 2018

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Fernandes, Fidel Henrique.

Gráficos de Controle para o Monitoramento de Processos Autorregressivo de Valores Inteiros com Inflação ou Deflação de Zeros / Fidel Henrique Fernandes. - 2018.

56 f.: il.

Dissertação (mestrado) - Universidade Federal do Rio Grande do Norte, Ciências Exatas e da Terra, Programa de Pós-Graduação em Matemática Aplicada e Estatística. Natal, RN, 2018.

Orientador: Prof. Dr. Marcelo Bourguignon Pereira.

1. Cadeia de Markov - Dissertação. 2. Deflação de zeros - Dissertação. 3. Gráfico CUSUM - Dissertação. 4. Gráfico de Shewhart - Dissertação. 5. Inflação de zeros - Dissertação. 6. Modelo ZMGINAR(1) - Dissertação. I. Pereira, Marcelo Bourguignon. II. Título.

RN/UF/BCZM

CDU 519.217.2

Dissertação de Mestrado sob o título *Gráficos de Controle para o Monitoramento de Processos Autorregressivo de Valores Inteiros com Inflação ou Deflação de Zeros* apresentada por Fidel Henrique Fernandes e aceita pelo Programa de Pós-Graduação em Matemática Aplicada e Estatística da Universidade Federal do Rio Grande do Norte, sendo aprovada por todos os membros da banca examinadora abaixo especificada:

Prof. Dr. MARCELO BOURGUIGNON PEREIRA
Orientador
Departamento de Estatística
UFRN

Prof^a. Dr^a. LUZ MILENA ZEA FERNANDEZ
Examinador Interno
Departamento de Estatística
UFRN

Prof^a. Dr^a. LINDA LEE HO
Examinador Externo
Departamento de Engenharia de Produção
USP

Natal-RN, 20 de Fevereiro de 2018

*Dedico este trabalho à minha mãe Adêjanes,
ao meu pai Flavianildo e meu irmão Fred.*

Agradecimentos

Agradeço a Deus, por ter me dado forças para continuar e lutar mesmos nos momentos mais difíceis.

Agradeço aos meus pais Flavianildo Henrique Fernandes e Maria Adêjanes Fernandes Henrique e ao meu irmão Fred pelo apoio, incentivo e preocupação com a minha Educação, e me fizeram perceber que o estudo necessário para a minha formação profissional, agradeço pela persistência em continuar no curso de Estatística, apesar de não ser de muito conhecimentos para eles.

Aos demais familiares que me ajudaram direta e indiretamente e me deram incentivo.

Aos amigos Daniel, Josenildo e Renato que conquistei durante os dois anos de convivência no mestrado, agradeço a companhia nos momentos de estudo e nos momentos de lazer. A Joyce por ter sido da minha turma desde a graduação, e a Wyara por ter te conhecido e pelo companheirismo.

A minha amiga Lilian por ter me encorajado nesses últimos anos e nunca ter duvidado da minha capacidade. Muito obrigado pelo seu apoio.

Ao Professor Dr. Pledson Guedes de Medeiros, por ter aceitado ser meu orientador durante a minha graduação e me ajudado desde o início do curso. Se não fosse pela persistência dele em melhorar o curso de Estatística, eu não estaria até hoje no curso. Espero que o senhor se recupere logo. Deus sempre no comando.

Ao professor Dr. Marcelo Bouguignon Pereira por ter sido o meu orientador durante essa jornada, agradeço demais o seu apoio, pela paciência, pelos seus conselhos, tornou-se não só o meu orientador, mas sim meu amigo.

A minha amiga Adrielle Oliveira por ter me ajudado na escrita do trabalho e por seus conselhos muito importantes e motivadores.

Aos Excelentes professores do PPGMAE, em especial à (Luz, Marcelo, André,

Roberto, Elaine e Fagner), um muito obrigado.

“A persistência é o caminho do êxito ”.

Charlie Chaplin

Gráficos de Controle para o Monitoramento de Processos Autorregressivo de Valores Inteiros com Inflação ou Deflação de Zeros

Autor: Fidel Henrique Fernandes

Orientador: Prof. Dr. Marcelo Bourguignon Pereira

Resumo

Uma série temporal é uma coleção de observações medidas sequencialmente ao longo do tempo, sendo estudadas com profunda notoriedade nos últimos anos juntamente com o controle estatístico de processo. O presente trabalho objetiva estudar o desempenho dos gráficos de controle CUSUM e Shewhart na detecção de médias do processo com inflação ou deflação de zeros através do modelo autorregressivo de valores inteiros geométrico zero modificado de primeira ordem [ZMGINAR(1)]. Nesse sentido, analisa-se ainda a sensibilidade através de simulações, o número médio de amostras que excedem o limite de controle (NMA), além disso, novos estimadores foram propostos afim de verificar, através do estudo de Monte Carlo, o comportamento dos estimadores através do erro quadrático médio (EQM) e do viés em diferentes cenários, os estimadores propostos se mostraram mais eficazes. No que concerne a simulação, os diferentes cenários apresentados com inflação e deflação de zeros, o CUSUM mostrou-se mais eficiente no cenário com deflação de zeros e o Shewhart com inflação de zeros em determinados casos. Nessa instância, considerou-se duas aplicações, uma com inflação de zeros e outra com deflação de zeros. Assim como na simulação, o CUSUM é melhor no cenário com deflação de zeros e o Shewhart com inflação de zeros. O grande diferencial deste trabalho é a aparição da deflação de zeros modelada nos gráficos de controle, além disso o modelo a ser trabalhado possui distribuição marginal conhecida diferentemente de outros modelos, o que é uma vantagem na implementação e construção de novos estimadores, acrescido a isso, considera-se ainda as estimativas dos parâmetros por diversos métodos: Máxima Verossimilhança, Yule-Walker e o estimador baseado em Probabilidade.

Palavras-chave: Cadeia de Markov; Controle Estatístico de Processos; deflação de zeros; gráfico CUSUM; gráfico de Shewhart; inflação de zeros; Modelo ZMGINAR(1); Monte Carlo; Número Médio de Amostras (NMA).

Control Charts for Monitoring of Autoregressive Processes of Integer Values with Inflation or Deflation of Zeros

Author: Fidel Henrique Fernandes

Advisor: Prof. Dr. Marcelo Bourguignon Pereira

Abstract

A time series is a collection of observations sequentially measured over time, being studied with great notoriety in recent years along with statistical process control. The present work aims to study the performance of the CUSUM and Shewhart control charts in the detection of averages of processes with inflation or deflation of zeros through the zero-modified geometric first-order integer-valued autoregressive process [ZMGINAR(1)]. In this sense, it is also analyzed the sensitivity through simulations, average number of samples exceeding the control limit (ARL), in addition, new estimators were proposed in order to verify, through the Monte Carlo study, the behavior of the estimators through the mean square errors (MSE) and of bias in different scenarios, the proposed estimators proved to be more effective. As for the simulation, the different scenarios presented with inflation and deflation of zeros, the CUSUM was more efficient in the scenario with zero deflation and Shewhart with zero inflation in certain cases. In this instance, two applications were considered, one with inflation of zeros and the other with deflation of zeros. Just as in the simulation, CUSUM is best in the zero deflation scenario and the Shewhart with zero inflation. The great differential of this work, is the appearance of deflation of zeros modeled in the control charts, in addition the model to be worked has a known marginal distribution unlike other models, which is an advantage in Implementation and construction of new estimators, in addition to this, it is still considered the parameter estimates by several methods: maximum likelihood, Yule-Walker and probability based estimator.

Keywords: average run length (ARL); CUSUM chart; deflation of zeros; inflation of zeros; Markov Chain; statistical process control; Monte Carlo; Shewhart Chart; ZMGINAR(1) model.

Lista de Figuras

2.1	Processo ZMGINAR(1) com $\mu = 0.5$ e $\alpha = 0.3$ e alguns valores de π com tamanho da amostra $n = 200$	11
2.2	Processo ZMGINAR(1) com $\mu = 5$ e $\alpha = 0.8$ e alguns valores de π com tamanho da amostra $n = 200$	12
2.3	Processo ZMGINAR(1) com $\pi = 0.1$ e $\alpha = 0.3$ e alguns valores de μ com tamanho da amostra $n = 200$	13
2.4	<i>Box plots</i> das estimativas dos parâmetros α, μ e π com tamanho da amostra $n = 400$. (a) $\alpha = 0.5$, (b) $\mu = 2$, (c) $\pi = -0.4$, (d) $\pi = -0.2$, (e) $\pi = 0.1$, (f) $\pi = 0.2$ e (g) $\pi = 0.3$	21
4.1	Gráficos da série, função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) para o número de pessoas que bebem em local público.	37
4.2	Gráficos de controle para o monitoramento do processo ZMGINAR(1) referentes ao número de pessoas que bebem em local público: (a) Gráfico de Shewhart (Fase 1: 1-100; Fase 2: 101-144), (b) Gráfico de CUSUM (Fase 1: 1-100; Fase 2: 101-144) e (c) Gráfico dos Saltos (Todas as observações).	40
4.3	Gráficos da série, função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) para o número de casos de Poliomelite.	41
4.4	Gráficos de controle para o monitoramento do processo ZMGINAR(1) referentes ao número casos de Poliomelite nos EUA: (a) Gráfico de Shewhart (Fase 1: 1-136; Fase 2: 136-167), (b) Gráfico de CUSUM (Fase 1: 1-136; Fase 2: 136-167) e (c) Gráfico dos Saltos (Todas as observações).	45

Lista de Tabelas

2.1	Estimativas médias e EQM (entre parênteses) provenientes dos resultados numéricos da simulação para valores verdadeiros de α, μ e π	20
3.1	<i>NMA</i> dos gráficos de Shewhart e CUSUM do processo ZMGINAR(1), $\alpha = 0.3$	30
3.2	<i>NMA</i> dos gráficos de Shewhart e CUSUM do processo ZMGINAR(1), $\alpha = 0.5$	31
3.3	<i>NMA</i> dos gráficos de Shewhart e CUSUM do processo ZMGINAR(1), $\alpha = 0.3$	32
3.4	<i>NMA</i> dos gráficos de Shewhart e CUSUM do processo ZMGINAR(1), $\alpha = 0.5$	34
4.1	Estatística descritiva do para o número de pessoas que bebem em local público.	37
4.2	Estimativa dos Parâmetros juntamente com seus respectivos erros padrão. .	38
4.3	Estatísticas descritivas para o número de pessoas que bebem em local público para as 100 primeiras observações.	38
4.4	Estimativa dos Parâmetros juntamente com seus respectivos erros padrão das 100 primeiras observações.	39
4.5	Estatísticas descritivas para o número de casos de Poliomelite.	42
4.6	Estimativa dos parâmetros com seus respectivos erros-padrão (entre parênteses), AIC e BIC.	42
4.7	Estatística Descritiva para o número de casos de Poliomelite das 136 primeiras obervaçãoes.	43
4.8	Estimativa dos parâmetros com seus respectivos erros-padrão (entre parênteses), AIC e BIC dos 136 primeiros casos de Poliomelite.	43

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	1
1.1	Objetivos da Dissertação	3
1.1.1	Objetivos Gerais	3
1.1.2	Objetivos Específicos	3
1.2	Organização da Dissertação	3
2	Processo Autorregressivo de Valores Inteiros com Inflação e Deflação de Zeros	5
2.1	Introdução	5
2.2	O processo ZMGINAR(1)	7
2.2.1	Distribuição de Zeros	12
2.2.2	Estimação dos parâmetros	14
2.2.2.1	Estimação por Mínimos Quadrados Condicionais (MQC)	14
2.2.2.2	Estimação por Máxima Verossimilhança	15
2.2.2.3	Estimação por Yule-Walker	16
2.2.2.4	Estimação baseada em Probabilidade	17
2.2.3	Estudo de Simulação de Monte Carlo	18
3	Gráficos de Controle para o processo ZMGINAR(1)	22
3.1	Gráficos de Shewart	23
3.2	Gráficos de CUSUM	24
3.3	Estudo Numérico	28

4	Aplicações	36
4.1	Inflação de Zeros	36
4.2	Deflação de Zeros	41
5	Considerações Finais	46

Capítulo 1

Introdução

O O Controle estatístico do processo (CEP) é uma ferramenta que tem por finalidade desenvolver e aplicar métodos estatísticos como parte de nossa estratégia para prevenção de defeitos, melhoria da qualidade de produtos e serviços e redução de custos. E tem como objetivo monitorar um produto ou serviço durante seu processo de produção.

O gráfico de controle é uma das técnicas principais do controle estatístico de processo, baseando-se em plotar medidas associada a alguma característica de qualidade do processo *versus* o tempo (ou o número de amostras), sendo uma ferramenta muito útil no monitoramento do processo. Quando fontes não-usuais de variabilidade estão presentes, os pontos serão plotados fora dos limites de controle, veja Montgomery (2007) para maiores detalhes. Em geral, o gráfico de controle é formado por uma linha central (LC) que representa o valor médio da estatística que está sendo monitorada, e por duas linhas externas denominadas limite superior de controle (LSC) e limite inferior de controle (LIC), denominados de modo a satisfazer algum critério de desempenho. Se a característica de qualidade considerada não puder ser conveniente apresentada numericamente, é comum na prática classificar cada item (ou unidade) inspecionado como conforme ou não-conforme, de acordo com as especificações.

Processos de dados de contagem surgem em muitas situações diferentes no CEP. Na Indústria de transformação, por exemplo, monitorar o número de defeitos ou não conformidades por unidade é de interesse em um processo de produção, enquanto na indústria de serviços monitorar o número de reclamações em um determinado período de tempo é uma característica importante da qualidade.

Em quase todas as áreas do conhecimento, existem fenômenos de interesse que se

desenvolvem e variam ao longo do tempo. As séries temporais de valores inteiros não-negativos, também referidas como séries de contagem, estão naturalmente presentes no nosso cotidiano. Como exemplos de aplicação, temos o número de elementos de uma população, o número de acidentes em uma determinada estrada ou o número de anomalias durante um certo período de tempo. No contexto de séries temporais de valores inteiros não-negativos de ordem 1, denominado INAR(1) os pioneiros foram McKenzie (1985) e Al-Osh e Alzaid (1987). Outros modelos de séries temporais de valores inteiros têm se destacado recentemente. Barreto-Souza (2015) introduziu o novo modelo INAR geométrico com marginal geométrica zero modificada e Bourguignon e Weiß (2017) introduziram um modelo de séries temporais de valores inteiros com subdispersão, sobredispersão e equidispersão. No contexto de CEP, Rakitzis, Weiß e Castagliola (2016) desenvolveram gráficos de controle para o monitoramento de processo de poisson com inflação de zeros, modelo este proposto por Jazi, Jones e Lai (2012). Li, Wang e Zhu (2016) desenvolveram gráficos de controle para o monitoramento do processo com distribuição marginal geométrica baseados no operador *thinning binomial negativo*, modelo este proposto por Ristic, Bakouch e Nastic (2009). Rakitzis, Castagliola e Maravelakis (2016) propuseram o gráfico de CUSUM para o monitoramento utilizando a distribuição Poisson inflada, Rakitzis, Maravelakis e Castagliola (2016) desenvolveram o gráfico de CUSUM para monitoramento do processo binomial inflado de zero e Rakitzis, Weiß e Castagliola (2017) introduziram gráficos de controle para monitorar contagens correlacionadas com um intervalo finito. Outros trabalhos desenvolvidos na literatura neste contexto são Weiß e Testik (2009), Weiß e Testik (2011), Weiß (2011), He, Huang e Woodall (2012), Rakitzis, Castagliola e Maravelakis (2016), Rakitzis e Castagliola (2016).

Seguindo esse contexto, são apresentados neste trabalho gráficos de controle para o monitoramento de séries temporais de valores inteiros com inflação ou deflação de zeros, sendo um modelo mais adequado para modelar esses tipos de cenários do que os modelos utilizados com distribuição Poisson, pois nem sempre é adequada para a modelagem, pois sua média e variância são as mesmas e essa propriedade pode ser inadequada para dados reais e além do que a vantagem do modelo é que o mesmo pode ser sobredisperso, subdisperso ou equidisperso. O grande diferencial deste trabalho ao que foi realizado em Rakitzis, Weiß e Castagliola (2016) é a inclusão da deflação de zeros modelada nos gráficos de controle, além disso o modelo a ser trabalhado possui distribuição marginal conhecida

diferentemente de outros modelos, o que é uma vantagem na implementação e construção de novos estimadores e também vamos considerar as estimativas dos parâmetros por diversos métodos: Máxima Verossimilhança, Yule-Walker e Probabilidade

1.1 Objetivos da Dissertação

1.1.1 Objetivos Gerais

O objetivo geral desta dissertação é propor e estudar o desempenho dos gráficos de controle CUSUM e Shewhart na detecção de desvios de médias do processo com inflação ou deflação de zeros através do modelo autorregressivo de valores inteiros geométrico zero modificado de primeira ordem [ZMGINAR(1)].

1.1.2 Objetivos Específicos

- Verificar o efeito da inflação e deflação de zeros, para diferentes combinações de parâmetros através dos gráficos de controle de Shewhart e CUSUM.
- Analisar o desempenho através de simulações do número médio de amostras necessárias até detectar mudança no processo (NMA), para diferentes mudanças na média submetidos a deflação ou inflação de zeros.
- Propor novos estimadores: Máxima Verossimilhança, Yule-Walker e o estimador baseado em Probabilidade para o modelo autorregressivo de valores inteiros com inflação ou deflação de zeros e verificar, através do estudo de Monte Carlo, o comportamento dos estimadores através do viés e EQM em diferentes cenários.
- Aplicar as metodologias estudadas para dois conjuntos de dados reais.

1.2 Organização da Dissertação

Esta dissertação é dividida em cinco capítulos. No segundo capítulo, descrevemos o modelo a ser estudado proposto por Barreto-Souza (2015) e verificamos o comportamento do modelo em diferentes cenários. Apresentamos um dos objetivos que é propor

novos estimadores. Nesse contexto, três métodos de estimação foram propostos para estimar os parâmetros do modelo e, ao final do capítulo, foram feitos estudos de simulações e encontramos que os estimadores propostos tiveram desempenho melhor em comparação aos estimadores propostos por Barreto-Souza (2015). Ressaltamos que alguns cálculos mostrados em Barreto-Souza (2015) precisaram ser retificadas durante o trabalho. No terceiro capítulo, apresentamos o outro objetivo do trabalho, que é a construção dos gráficos de controle de Shewhart e CUSUM na detecção de médias do processo com inflação ou deflação de zeros através do modelo ZMGINAR(1) e ao final do capítulo realizou-se estudos numéricos. No quarto capítulo, são apresentados duas aplicações: uma com inflação de zeros e outra com deflação de zeros. Por fim, no quinto capítulo, são apresentadas conclusões e comentários.

Capítulo 2

Processo Autorregressivo de Valores Inteiros com Inflação e Deflação de Zeros

O objetivo deste capítulo é apresentar o modelo autorregressivo de valores inteiros geométrico zero modificado de primeira ordem ZMGINAR(1) com inflação e deflação de zeros proposto por Barreto-Souza (2015), bem como estudar suas propriedades, além de propor estimadores e compará-los com os já existentes: Máxima Verossimilhança, Yule-Walker e o estimador baseado em Probabilidade e, por fim, avaliamos através de um estudo de simulação de Monte Carlo. Algumas demonstrações podem ser encontradas no Apêndice.

2.1 Introdução

Em quase todas as áreas do conhecimento, existem fenômenos de interesse que se desenvolvem e variam ao longo do tempo. Neste tipo de dados é comum existir uma estrutura de correlação entre as observações. Essas estruturas de dados que se desenvolvem espaçadamente no tempo, são denominados séries temporais.

Uma série temporal é a realização de um processo estocástico, em que T é o domínio do tempo e X_t é uma variável aleatória, tal que $t \in T$. Um processo estocástico é uma família de variáveis aleatórias representando a evolução de um sistema de valores com o tempo. Tecnicamente, o domínio do tempo T pode ser discreto ou contínuo. Uma

série temporal é dita ser contínua quando as observações são feitas continuamente no tempo. Uma série temporal é discreta se o conjunto dos tempos em que as observações são realizadas, T , é discreto. Neste trabalho consideramos a série temporal discreta, isto é, $T = \mathbb{Z}$. Se as variáveis aleatórias que compõem o processo seguirem uma distribuição discreta, o processo é dito de marginal discreta e é conhecido como processo de contagem ou processo de valores inteiros.

Séries temporais de valores inteiros (processos de contagem) têm sido consideradas em vários artigos, principalmente por causa das áreas de aplicação. Com a necessidade de modelar, por exemplo, o número de sucessos de um determinado fenômeno é notório admitir que a variável aleatória associada a este fenômeno segue uma distribuição discreta. Os modelos lineares para séries temporais não são adequados para modelar processos de contagem, pois o produto de uma constante real por uma variável aleatória de valor inteiro pode produzir uma variável aleatória real (PEREIRA, 2011). Al-Osh e Alzaid (1987) recorreram à operação *thinning binomial* introduzida por Steutel e Harn (1979) para substituir a operação de multiplicação usual e propuseram o modelo autoregressivo de valores Inteiros (INAR).

Inúmeras pesquisas sobre os modelos de séries temporais de valores inteiros vêm ganhando espaço no meio acadêmico há mais de três décadas. Johansson (1996) utilizou séries temporais de contagem na área de prevenção de acidentes. Já na área da saúde pode ser útil para modelar o número de infectados por uma determinada doença ao longo do tempo, ver Ferland, Latour e Oraichi (2006). Gomes (2009) usou o modelo $\text{INAR}(p)$ para previsão de índices de qualidade de ar. Silva (2016) introduziu a previsão e a estimação do modelo $\text{INARCH}(2)$. Lopes (2016) introduziu novos modelos de séries temporais binárias de contagem baseadas em operadores *thinning*.

Séries temporais de valores inteiros são obtidas através de processos de contagem definidos como o número de acontecimentos, objetos ou indivíduos em intervalos consecutivos de tempo. Como exemplos tem-se o número de nascimentos ou mortes em um hospital a cada dia, o número de acidentes numa fábrica a cada mês, o número de viajantes de comboio a cada hora entre inúmeras outras possibilidades.

Diversos trabalhos podem ser citados nas mais diferentes áreas das ciências tais como sistemas de fila de espera Ahn, Lee e Jeon (2000), biologia experimental Zhou e Basawa (2005), epidemiologia Zheng e Basawa (2008) e telecomunicações Weiß (2008).

2.2 O processo ZMGINAR(1)

Seja X uma variável aleatória seguindo uma distribuição geométrica zero modificada (ZMG) com parâmetros $\mu > 0$ e $\pi \in (-1/\mu, 1)$. Dessa forma, X tem a seguinte função de probabilidade

$$p(k) = \mathbb{P}(X = k) = \begin{cases} \pi + (1 - \pi)\frac{1}{1+\mu}, & k = 0, \\ (1 - \pi)\frac{\mu^k}{(1+\mu)^{k+1}}, & k = 1, 2, \dots \end{cases} \quad (2.1)$$

Neste trabalho, vamos denotar $X \sim \text{ZMG}(\pi, \mu)$. Esta distribuição apresenta algumas particularidades envolvendo seus parâmetros. Segundo Barreto-Souza (2015) para $\pi \in (0, 1)$, temos que $X \stackrel{d}{=} YZ$ ¹, em que Y e Z são variáveis aleatórias independentes seguindo uma distribuição Bernoulli e geométrica, respectivamente. Para $\pi \in (-1/\mu, 0)$ e $\pi \in (0, 1)$, temos um modelo com deflação de zeros e um modelo com inflação de zeros, respectivamente. A inflação de zeros significa, que há mais zeros do que esperado, já a deflação de zeros refere-se a menos zeros do que era esperado. A distribuição ZMG tem a distribuição geométrica como um caso especial quando $\pi = 0$. A função geradora de probabilidade (f.g.p) de X , $\varphi_X(s)$, definida por $\varphi_X(s) = E(s^X)$ é dada por

$$\varphi_X(s) = \frac{1 + \pi\mu(1 - s)}{1 + \mu(1 - s)},$$

para $|s| < (1 + \mu)/\mu$. A média e a variância de X são dados por

$$E(X) = \mu(1 - \pi) \quad (2.2)$$

e

$$\text{Var}(X) = \mu(1 - \pi)[1 + \mu(1 + \pi)].$$

Outra propriedade importante é o índice de dispersão, definido como $I_X = \text{Var}(X)/E(X) = 1 + \mu(1 + \pi)$. Se $\mu \in (0, 1)$ e $\pi \in (-1/\mu, 1)$, a distribuição ZMG apresenta subdispersão. Se $\mu \in (0, 1)$ e $\pi = -1$, a distribuição ZMG apresenta equidispersão e para $\pi \in (-1, 1)$ a distribuição ZMG apresenta sobredispersão. Antes de definir o processo ZMGINAR(1), vamos definir abaixo o operador *thinning* binomial negativo.

¹O seguinte simbolo $\stackrel{d}{=}$ denota que as variáveis aleatórias da igualdade possuem a mesma distribuição.

Definição 2.2.1. *Seja X uma variável aleatória de valor inteiro não-negativo e α um número real tal que $\alpha \in [0, 1)$. O operador thinning binomial negativo, denotado por “ $*$ ”, é definido por Ristic, Bakouch e Nastic (2009) como*

$$\alpha * X = \sum_{i=1}^X G_i, \quad (2.3)$$

para $X > 0$ e $\alpha * 0 = 0$, em que $\{G_i\}_{i \geq 1}^X$ é uma sequência de v.a. i.i.d. com distribuição geométrica com média α . Mais especificamente, G_i tem função de probabilidade dada em (2.1) quando $\pi = 0$ e $\mu = \alpha$, para $i \geq 1$. Observe que $\alpha * X|X = x$, tem distribuição binomial negativa de parâmetros α e x .

Em decorrência da definição do operador *thinning* binomial negativo, temos o seguinte lema.

Lema 2.2.1. *Seja X uma variável aleatória de valor inteiro não-negativo se as séries de contagem de $\alpha_i * X_i$, $i = 1, 2, \dots, r$, são mutuamente independentes de X_i , $i = 1, 2, \dots, r$. O operador thinning binomial negativo “ $*$ ” apresentado em (2.3) tem as seguintes propriedades*

- i) $E[\alpha * X] = \alpha E[X]$;
- ii) $E[\alpha * X|X] = \alpha X$;
- iii) $Var[\alpha * X|X] = \alpha(1 + \alpha)X$;
- iv) $Var[\alpha * X] = \alpha^2 Var[X] + \alpha(1 + \alpha)E[X]$;
- v) $E[\alpha * X]^2 = \alpha(1 + \alpha)E[X] + \alpha^2 E[X^2]$;
- vi) $E[\alpha * X]^3 = \alpha E[X^3] + 3\alpha^2(1 + \alpha)E[X^2] + \alpha(1 + \alpha)(1 + 2\alpha)E[X]$;
- vii) $E\left[\prod_{i=1}^r (\alpha_i * X_i)\right] = \prod_{i=1}^r \alpha_i E\left[\prod_{i=1}^r X_i\right], r \geq 1$.

A prova dessas propriedades estão disponíveis no Apêndice. A prova é semelhante ao que foi apresentado por Silva e Oliveira (2004) e alguns itens são demonstrados em Ristic, Bakouch e Nastic (2009). Barreto-Souza (2015) mostra duas proposições que são importantes para a definição do processo ZMGINAR(1).

Proposição 2.2.1. *Seja $X \sim ZMG(\pi, \mu)$ e “ $*$ ” o operador thinning binomial negativo definido em (2.3). Então $\alpha * X \sim ZMG\left(\frac{1+\pi\mu}{1+\mu}, \alpha(1+\mu)\right)$, para $\alpha > 0$.*

Proposição 2.2.2. *Seja $\phi(s) = \varphi_X(s)/\varphi_{\alpha*X}(s)$, em que $\varphi_X(s)$ e $\varphi_{\alpha*X}(s)$ são as f.g.p de X e $\alpha * X$, respectivamente. Para $\alpha \in \left(\max\left\{0, \frac{\pi\mu}{1+\pi\mu}\right\}, \frac{\mu}{1+\mu}\right)$, temos que $\phi(s)$ é a f.g.p da variável aleatória $Z \stackrel{d}{=} Z_1 + Z_2$, em que Z_1 e Z_2 são duas variáveis aleatórias independentes com $Z_1 \sim ZMG\left(\alpha\frac{1+\mu}{\mu}, \mu\right)$ e $Z_2 \sim ZMG\left(\frac{\pi\mu\alpha^{-1}}{1+\pi\mu}, \alpha(1+\pi\mu)\right)$.*

As provas das Proposições 2.2.1 e 2.2.2 são apresentadas por Barreto-Souza (2015).

Definição 2.2.2. *Seja $\{X_t\}_{t \geq 0}$ uma sequência estacionária com distribuição marginal $ZMG(\pi, \mu)$. O processo ZMGINAR(1) é definido como*

$$X_t = \alpha * X_{t-1} + \epsilon_t, \quad (2.4)$$

para $t \geq 0$, ou seja, os números naturais sem o zero, em que $\{\epsilon_t\}_{t \geq 0}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas independente da série de contagem e X_{t-l} e ϵ_t são independentes para todo $l \geq 1$.

Pela Proposição 2.2.2 temos que a distribuição das inovações é uma convolução entre duas variáveis aleatórias independentes com distribuição ZMG. Mais especificamente, temos que $\epsilon_t \stackrel{d}{=} Z_1 + Z_2$, com Z_1 e Z_2 independentes, sendo a distribuição de $Z_1 \sim ZMG\left(\alpha\frac{1+\mu}{\mu}, \mu\right)$ e $Z_2 \sim ZMG\left(\frac{\pi\mu\alpha^{-1}}{1+\pi\mu}, \alpha(1+\pi\mu)\right)$. Assim a distribuição de probabilidade de ϵ_t , denotada por $p_\epsilon(\cdot)$, é dada por

$$p_\epsilon(0) \equiv \mathbb{P}(\epsilon_t) = \frac{1 + \pi\mu}{1 + \alpha(1 + \pi\mu)} \left(\alpha + \frac{1}{1 + \mu} \right) \quad (2.5)$$

e

$$\begin{aligned} p_\epsilon(k) \equiv \mathbb{P}(\epsilon_t = k) &= \frac{(1 + \pi\mu)[\mu - \alpha(1 + \mu)]\mu^{k-1}}{[1 + \alpha(1 + \pi\mu)](1 + \mu)^{k+1}} + \\ &\frac{[1 + \alpha(1 + \mu)][\alpha(1 + \pi\mu) - \pi\mu](1 + \pi\mu)^{k-1}\alpha^{k-1}}{(1 + \mu)[1 + \alpha(1 + \pi\mu)]^{k+1}} + \\ &\frac{[\alpha(1 + \pi\mu) - \pi\mu][\mu - \alpha(1 + \mu)]\mu^k}{\alpha\mu(1 + \pi\mu)[1 + \alpha(1 + \pi\mu)](1 + \mu)^{k+1}} \times \\ &\frac{\mu[1 + \alpha(1 + \pi\mu)] \left[\frac{\alpha(1+\mu)(1+\pi\mu)}{\mu(1+\alpha(1+\pi\mu))} \right]^k - \alpha(1 + \mu)(1 + \pi\mu)}{\alpha(1 + \pi\mu) - \mu}, \end{aligned}$$

para $k \geq 1$, sendo k um número inteiro. A f.g.p de ϵ_t , denotado por $\varphi_\epsilon(s)$, é dada por $\phi(s)$ definido na Proposição 2.2.2. A média e a variância de ϵ_t são dadas, respectivamente, por $E(\epsilon_t) = \mu(1-\pi)(1-\alpha)$ e $Var(\epsilon_t) = \mu(1-\pi)\{(1-\alpha)[1+\mu(1+\pi)(1+\alpha)]-2\alpha^2\}$. O processo ZMGINAR(1) apresentado em (2.4) está definido apenas para $\alpha \in \left(\max\left\{0, \frac{\pi\mu}{1+\pi\mu}\right\}, \frac{\mu}{1+\mu}\right)$ (condição de estacionariedade). Segundo Barreto-Souza (2015) o ZMGINAR(1) é um processo markoviano, estacionário e ergódico. Podemos observar o comportamento das trajetórias simuladas do processo ZMGINAR(1) para diferentes valores dos parâmetros apresentados abaixo nas Figuras 2.1, 2.2 e 2.3.

A Figura 2.1, apresenta as trajetórias para os valores de $\mu = 0.5$ e $\alpha = 0.3$ e alguns valores negativos de π , incluindo o zero. E a Figura 2.2, apresenta as trajetórias para os valores de $\mu = 5$ e $\alpha = 0.8$ e alguns valores positivos de π , incluindo o zero. Vale ressaltar que, quando $\pi = 0$ corresponde ao modelo NGINAR(1) proposto por Ristic, Bakouch e Nastic (2009). Conforme Barreto-Souza (2015), nota-se que a quantidade de zeros tende a aumentar à medida que o π aumenta, essa propriedade é importante, pois não observa-se esse comportamento no processo NGINAR(1), visto que é um caso particular do modelo ZMGINAR(1). Na Figura 2.3, observam-se séries de dados com inflação de zeros e a medida que o μ aumenta a proporção de zeros diminui. As probabilidades de transição de um passo, especificamente, conhecidas como probabilidades condicionais, que são definidas como a probabilidade do processo X_t estar no tempo t no estado j , $X_t = j$, dado que no tempo $t - 1$ estava no estado i , denotado por $p_{j|i}$. Já o $p_{j|0}$, denota a probabilidade do processo X_t estar no tempo t no estado j dado que estava no estado $i = 0$ no tempo $t - 1$. As probabilidades de transição são particularmente úteis para obter estimativas para os parâmetros do modelo, através do método de máxima verossimilhança. Portanto, as probabilidades de transição do modelo ZMGINAR(1) dado em (2.4), utilizando as propriedades do operador *thinning* binomial negativo são dadas por:

$$p_{j|0} = p_\epsilon(j) \quad \text{e} \quad p_{j|i} = \sum_{k=0}^j \binom{k+i-1}{k} \alpha^k (1+\alpha)^{-(k+i)} p_\epsilon(j-k), \quad i \geq 1, \quad (2.6)$$

em que $p_\epsilon(\cdot)$ é a função de probabilidade das inovações dada em (2.5). É pertinente salientar, que os cálculos apresentados por Barreto-Souza (2015), mostraram necessidades de retificações, estas que estão explícitas no presente trabalho, que foram as probabilidades

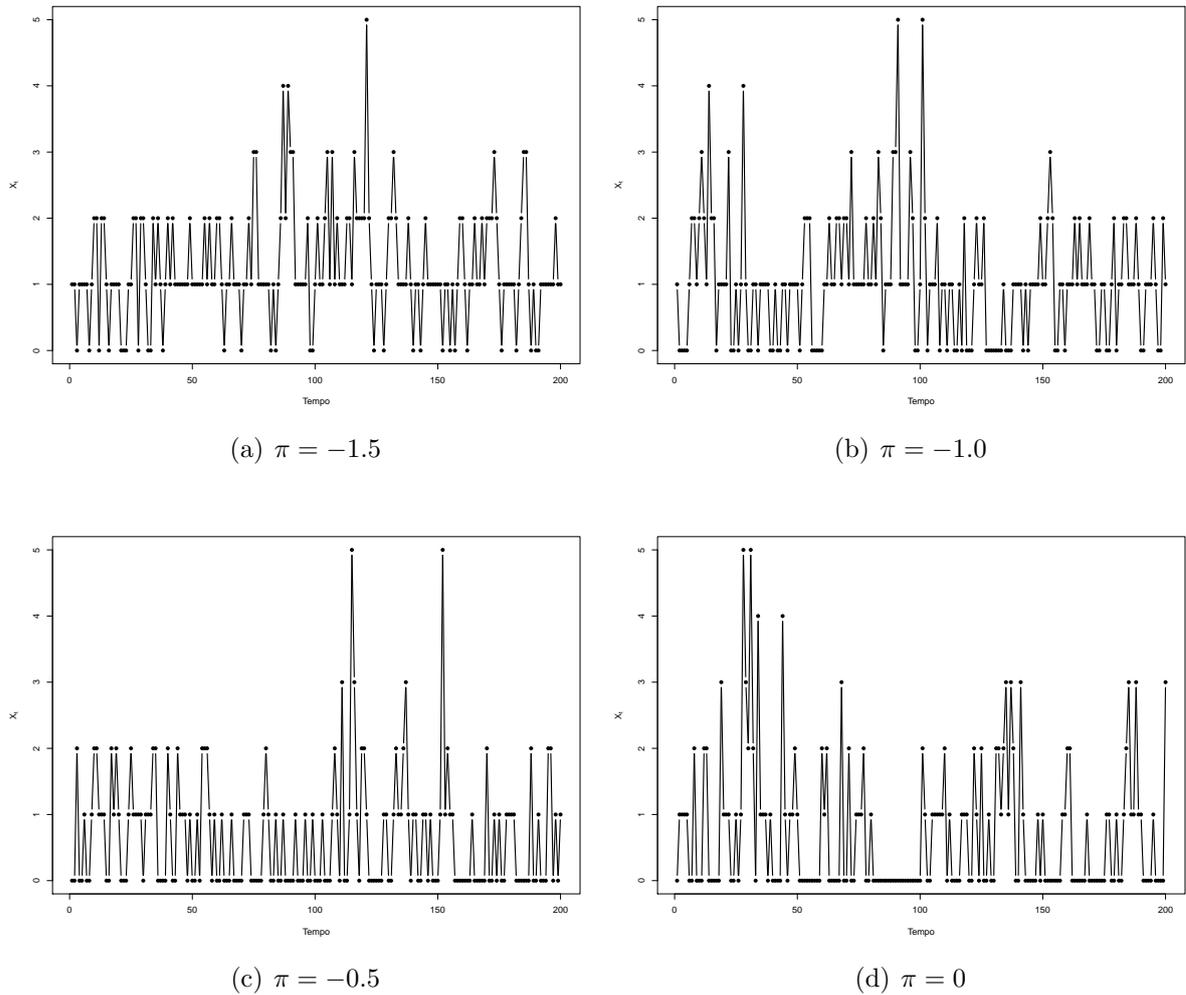


Figura 2.1: Processo ZMGINAR(1) com $\mu = 0.5$ e $\alpha = 0.3$ e alguns valores de π com tamanho da amostra $n = 200$.

Fonte: Elaborada pelo autor

de transição e a função de autocorrelação dos saltos. Como $\alpha * X_{t-1} | X_{t-1} = j$ segue uma distribuição Binomial Negativa com parâmetros α e j , assim, o valor $k \geq 0$ e ϵ_t segue uma distribuição dada em 2.5, assim, $j - k \geq 0 \Rightarrow k \leq j$. Portanto, o somatório $\sum_{k=0}^j$ na probabilidade de transição $p_{i|j}$ trata-se de uma convolução entre a binomial negativa e a distribuição das inovações, assim, o máximo que pode atingir é o valor j .

Teorema 2.2.1. *Seja $\{X_t\}_{t \geq 0}$ um processo estacionário ZMGINAR(1) e $\varphi_\epsilon(v)$ a função geradora de probabilidade da variável ϵ_t , então a função geradora de probabilidade conjunta*

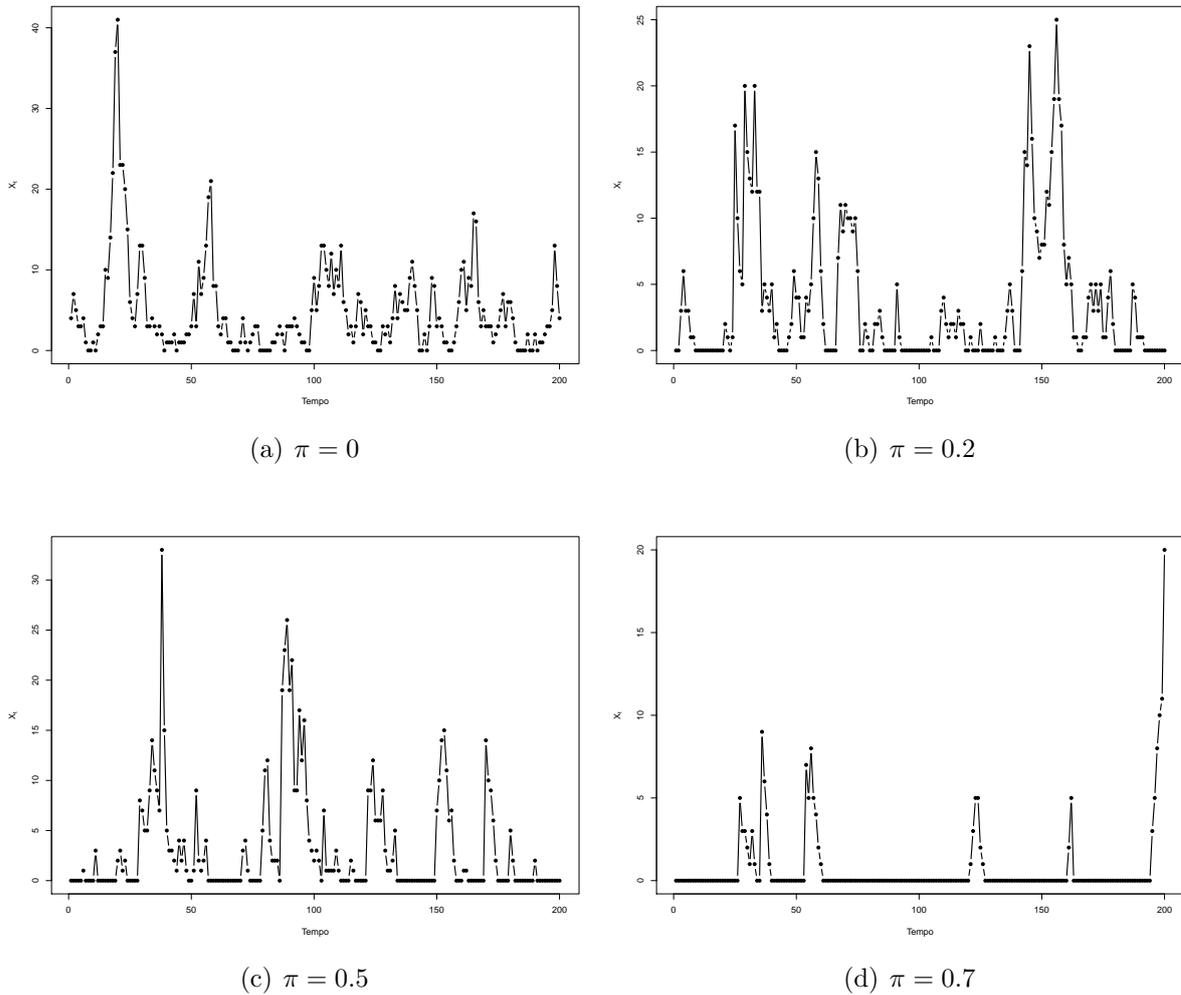


Figura 2.2: Processo ZMGINAR(1) com $\mu = 5$ e $\alpha = 0.8$ e alguns valores de π com tamanho da amostra $n = 200$.

Fonte: Elaborada pelo autor

de X_t e X_{t-1} é dada por:

$$\varphi_{X_{t-1}, X_t}(u, v) = \varphi_\epsilon(v) \frac{(1 + \pi\mu)[1 + \alpha(1 - v)] - \pi\mu u}{(1 + \mu)[1 + \alpha(1 - v)] - \mu u}.$$

A demonstração do Teorema 2.2.1 pode ser consultada no Apêndice.

2.2.1 Distribuição de Zeros

Para explicar sobre a quantidade de zeros que aparecem nas Figuras 2.1 e 2.2 iremos abordar a distribuição de zeros. Para o processo ZMGINAR(1), a probabilidade

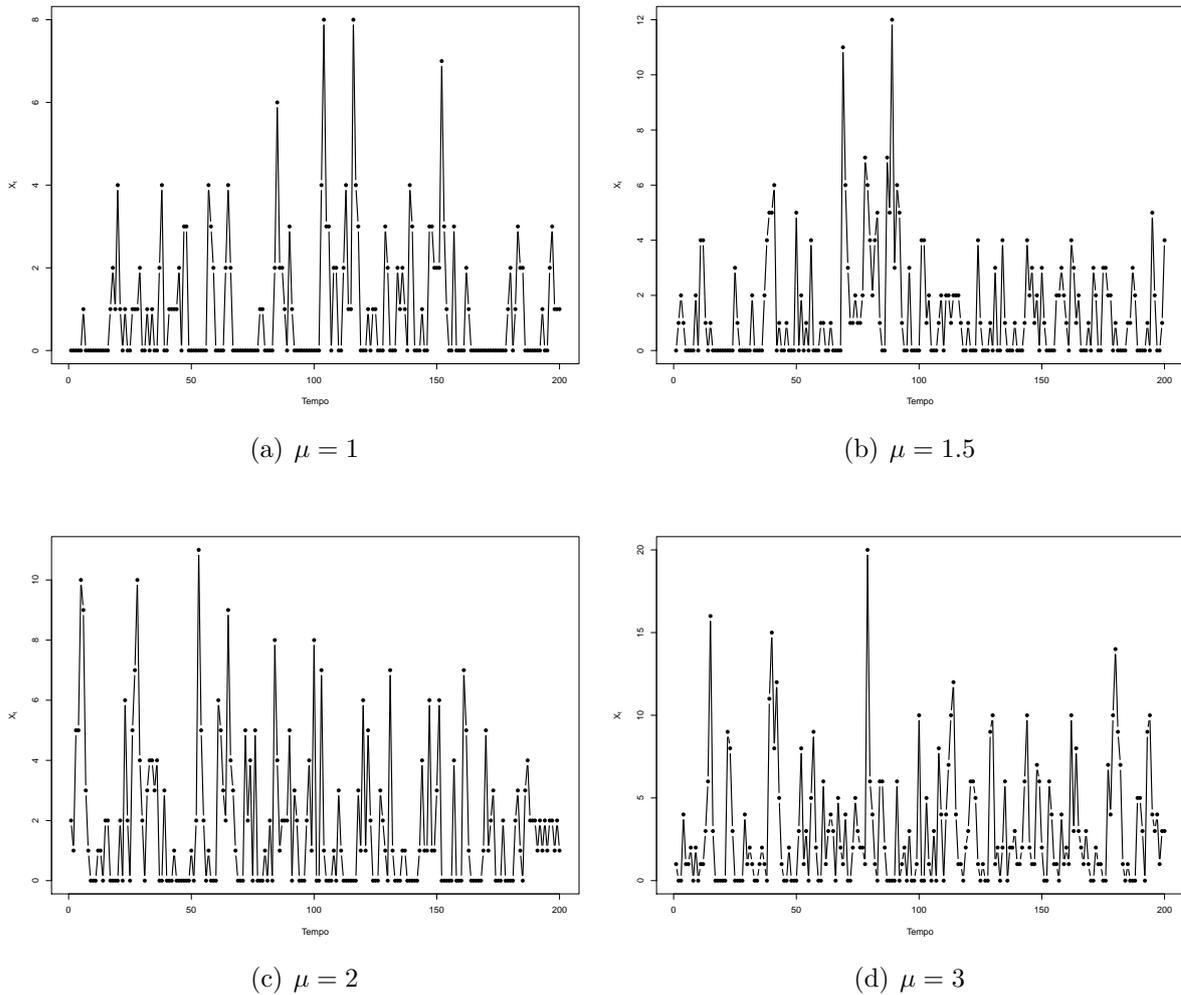


Figura 2.3: Processo ZMGINAR(1) com $\pi = 0.1$ e $\alpha = 0.3$ e alguns valores de μ com tamanho da amostra $n = 200$.

Fonte: Elaborada pelo autor

de transição de zero para zero é a distribuição das inovações $p_\epsilon(0)$, dada em (2.5). Com isso, Barreto-Souza (2015) obteve a probabilidade de transição de zero para não-zero, denotada por δ ,

$$\delta = \mathbb{P}(X_t \neq 0 | X_{t-1} = 0) = \frac{(1 - \pi)\mu}{(1 + \mu)[1 + \alpha(1 + \pi\mu)]}. \quad (2.7)$$

Proposição 2.2.1.1. *O comprimento da sequência de zeros segue uma distribuição geométrica com a função de probabilidade dada por $\mathbb{P}(N = k) = \delta(1 - \delta)^{k-1}$, com $k \in \mathbb{N}^*$, em que δ é dado em (2.7). Em particular, o valor esperado do comprimento da sequência*

de zeros é dado por

$$E(N) = \frac{(1 + \mu)[1 + \alpha(1 + \pi\mu)]}{(1 - \pi)\mu}. \quad (2.8)$$

A prova da Proposição 2.2.1.1 se encontra em Barreto-Souza (2015). Note que, o valor esperado do comprimento de zeros para o processo NGINAR(1) é obtido fazendo $\pi = 0$ em (2.8), ou seja, $E_0(N) = (1 + \mu)(1 + \alpha)/\mu$. Podemos verificar que $E(N) = (1 - \pi)^{-1}[E_0(N) + \alpha\pi(1 + \mu)]$ e nota-se que $E(N) \geq E_0(N)$ para $\pi \in [0, 1)$ e $E(N) \leq E_0(N)$ quando $\pi \in (-1/\mu, 0)$. Verifica-se também que $E(N)$ é uma função crescente em π , dessa forma, esclarece a presença da quantidade de zeros nas Figuras 2.1 e 2.2.

2.2.2 Estimação dos parâmetros

Barreto-Souza (2015) utilizou o estimador de mínimos quadrados condicionais (MQC) para estimar os parâmetros do modelo ZMGINAR(1). Neste trabalho, apresentamos novos estimadores: Estimador de Máxima Verossimilhança (MV), Yule-Walker (YW) e o Estimador baseado em Probabilidade (PB).

2.2.2.1 Estimação por Mínimos Quadrados Condicionais (MQC)

Seja $\boldsymbol{\theta} = (\pi, \mu, \alpha)^\top$ o vetor de parâmetros e X_1, \dots, X_n uma amostra de tamanho n . Seja $Q_n(\boldsymbol{\theta}) = \sum_{t=2}^n [X_t - E(X_t|X_{t-1})]^2$ a função que deseja-se minimizar, e como $E(X_t|X_{t-1}) = \alpha X_{t-1} + \mu(1 - \pi)(1 - \alpha)$, temos que, $Q_n(\boldsymbol{\theta}) = \sum_{t=2}^n [X_t - \mu(1 - \pi)(1 - \alpha) - \alpha X_{t-1}]^2$, logo, os argumentos de $\boldsymbol{\theta}$ que minimizam $Q_n(\boldsymbol{\theta})$ são os estimadores de mínimos quadrados condicionais. Para obtê-los é necessário encontrar a solução do sistema de equações $\partial Q_n(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = 0$. Assumindo μ conhecido e fazendo $\partial Q_n(\boldsymbol{\theta})/\partial(\pi, \alpha)^\top = 0$, resolvendo este sistema de equações, chega-se aos seguintes estimadores de mínimos quadrados condicionais

$$\hat{\alpha}_{MQC} = \frac{(n-1) \sum_{t=2}^n X_t X_{t-1} - \sum_{t=2}^n X_t \sum_{t=2}^n X_{t-1}}{(n-1) \sum_{t=2}^n X_{t-1}^2 - (\sum_{t=2}^n X_{t-1})^2}$$

e

$$\hat{\pi}_{MQC} = 1 - \frac{\sum_{t=2}^n X_t - \hat{\alpha}_{MQC} \sum_{t=2}^n X_{t-1}}{\mu(n-1)(1 - \hat{\alpha}_{MQC})}.$$

Proposição 2.2.2.1.1. *Seja π, μ e α os valores verdadeiros dos parâmetros. Os estimadores $\hat{\pi}_{MQC}$ e $\hat{\alpha}_{MQC}$ são estimadores consistentes para π e α , respectivamente, segue a distribuição assintótica dos estimadores.*

$$\sqrt{n}\{(\hat{\alpha}_{MQC}, \hat{\pi}_{MQC}) - (\alpha, \pi)\} \xrightarrow{d} N((0, 0), \Sigma),$$

em que

$$\Sigma = \begin{pmatrix} \sigma_{\alpha}^2 & \sigma_{\alpha\pi} \\ \sigma_{\alpha\pi} & \sigma_{\pi}^2 \end{pmatrix}$$

com

$$\sigma_{\alpha}^2 = \frac{\mu^3\{[1 + (1 - \pi)(\pi - \alpha(1 + \pi)^2)]\} + 2\mu^2[1 + \pi(\alpha + \pi(2\alpha - 1))] + \mu[\pi(4\alpha - 1) + 2\alpha + 1] + \alpha}{\mu(1 + \alpha)^{-1}(1 - \pi)[1 + \mu(1 + \pi)]^2},$$

$$\sigma_{\alpha\pi} = -\frac{\alpha(1 + \alpha)}{\mu(1 - \alpha)} \quad \text{e} \quad \sigma_{\pi}^2 = \frac{(1 + \alpha)(1 - \pi)[1 + \mu(1 + \pi)]}{\mu(1 - \alpha)}.$$

A prova da Proposição 2.2.2.1.1 se encontra em Barreto-Souza (2015).

A normalidade assintótica de $\sqrt{n}(\hat{\pi} - \pi)$ é de interesse particular, pois podemos testar as hipóteses $H_0 : \pi = 0$ contra as hipóteses alternativas $H_1 : \pi < 0$ e $H_2 : \pi > 0$. No presente trabalho, estamos interessados em testar modelo NGINAR(1) contra o modelo ZMGINAR(1) na presença de inflação ou deflação de zeros. Barreto-Souza (2015) propôs o seguinte teste. Seja q_{γ} o quantil da distribuição normal padrão, ou seja, $\Phi(q_{\gamma}) = \gamma$, para $\gamma \in (0, 1)$, em que, $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão. Rejeita-se a hipótese nula $H_0 : \pi = 0$ em favor da hipótese alternativa $H_1 : \pi < 0$ para deflação de zeros se

$$\hat{\pi} < q_{\gamma} \sqrt{\frac{(1 + \hat{\alpha})(1 + \hat{\mu})}{n\hat{\mu}(1 - \hat{\alpha})}},$$

em que γ é o nível de significância do teste e $\hat{\mu}$ é qualquer estimador consistente. E se a hipótese alternativa for $H_1 : \pi > 0$ para a inflação de zeros, a região crítica é dada por

$$\hat{\pi} > q_{1-\gamma} \sqrt{\frac{(1 + \hat{\alpha})(1 + \hat{\mu})}{n\hat{\mu}(1 - \hat{\alpha})}}.$$

2.2.2.2 Estimação por Máxima Verossimilhança

Seja X_1, X_2, \dots, X_n uma amostra de tamanho n do modelo ZMGINAR(1). Sabendo que $X_1 \sim ZMG(\pi, \mu)$ e a distribuição condicional de X_t dado X_{t-1} apresentada em (2.6),

temos que a função de log-verossimilhança para o processo ZMGINAR(1) é dada por

$$\ell(\alpha, \mu, \pi) = \log[\mathbb{P}(X_1 = x_1)] + \sum_{t=2}^n \log[\mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1})].$$

Os valores de α , μ e π que maximizam a função de log-verossimilhança $\ell(\alpha, \mu, \pi)$ são os mesmos que maximizam a função de verossimilhança, denotados por $\hat{\alpha}_{MV}$, $\hat{\mu}_{MV}$ e $\hat{\pi}_{MV}$. Os estimadores de máxima verossimilhança não têm forma fechada e a maneira de se obter os estimadores é através de métodos numéricos usando a otimização não linear, como por exemplo através da função `optim` do software R Core Team (2013).

2.2.2.3 Estimação por Yule-Walker

O estimador de Yule-Walker é obtido utilizando as equações de Yule-Walker e também é conhecido como estimador pelo método dos momentos, uma vez que estima os parâmetros por meio das correlações amostrais e também pelos momentos amostrais do modelo. Os parâmetros α , π e μ serão estimados, utilizando a primeira autocorrelação amostral, o primeiro momento e o segundo momento, respectivamente. Como a $\rho(k) = \alpha^k$ o estimador de α será dado por $\hat{\alpha}_{YW} = \hat{\rho}(1)$, assim

$$\hat{\alpha}_{YW} = \frac{\sum_{t=2}^n (X_t - \bar{X})(X_{t-1} - \bar{X})}{\sum_{t=2}^n (X_t - \bar{X})^2},$$

em que X_1, X_2, \dots, X_n é uma amostra do processo $\{X_t\}_{t \in \mathbb{Z}}$ e $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$. O primeiro momento é dado por

$$E(X_t) = \mu(1 - \pi) \tag{2.9}$$

Dessa forma, igualando o primeiro momento amostral com o primeiro momento populacional dada em (2.9), tem-se

$$\frac{1}{n} \sum_{t=1}^n X_i = \mu(1 - \pi) \Rightarrow \bar{X} = \hat{\mu}(1 - \hat{\pi}) \Rightarrow \hat{\pi}_{YW} = 1 - \frac{\bar{X}}{\hat{\mu}_{YW}}.$$

O segundo momento é dado por

$$\begin{aligned} E(X_t^2) &= \text{Var}(X_t) + [E(X_t)]^2 = \mu(1 - \pi)[1 + \mu(1 + \pi)] + \mu^2(1 - \pi)^2 \\ &= \mu(1 - \pi)(2\mu + 1), \end{aligned} \quad (2.10)$$

e igualando a expressão (2.10) ao segundo momento amostral e fazendo as devidas substituições dos parâmetros μ e π pelos seus respectivos estimadores, tem-se

$$\frac{1}{n} \sum_{t=1}^n X_t^2 = \mu(1 - \pi)(2\mu + 1).$$

Dessa forma, o estimador de μ é dado por

$$\hat{\mu}_{YW} = \frac{1}{2} \left(\frac{\sum_{t=1}^n X_t^2}{\sum_{t=1}^n X_t} - 1 \right).$$

Proposição 2.2.2.3.1. $\hat{\alpha}_{YW}$, $\hat{\mu}_{YW}$ e $\hat{\pi}_{YW}$ são estimadores consistentes para estimar α , μ e π , respectivamente.

Demonstração. De acordo com Bourguignon e Weiß (2017) a consistência desses estimadores resulta da ergodicidade do processo e as propriedades de convergência em probabilidade. \square

2.2.2.4 Estimação baseada em Probabilidade

Os estimadores baseado em Probabilidade são uma alternativa para obtermos novos estimadores. A partir da Equação (2.1), obtem-se $\mathbb{P}(X_t > 0) = \frac{\mu(1-\pi)}{1+\mu}$, $\mathbb{P}(X_t > 1) = \frac{\mu^2(1-\pi)}{(1+\mu)^2}$ e $\mathbb{P}(X_t = 1) = \frac{\mu(1-\pi)}{(1+\mu)^2}$. Dessa forma, fazendo $\frac{\mathbb{P}(X_t > 1)}{\mathbb{P}(X_t = 1)} = \mu$ e $1 - \frac{[\mathbb{P}(X_t > 0)]^2}{\mathbb{P}(X_t > 1)} = \pi$, implica que os estimadores baseados em probabilidade (PB) dos parâmetros μ e π são dados por

$$\hat{\mu}_{PB} = \frac{\sum_{t=1}^n \mathbb{I}_{\{X_t > 1\}}}{\sum_{t=1}^n \mathbb{I}_{\{X_t = 1\}}} \quad \text{e} \quad \hat{\pi}_{PB} = 1 - \frac{\left(\sum_{t=1}^n \mathbb{I}_{\{X_t > 0\}} \right)^2}{\sum_{t=1}^n \mathbb{I}_{\{X_t > 1\}}},$$

em que $\mathbb{I}_{\{\cdot\}}$ é a função indicadora. O parâmetro α pode ser estimado usando a probabilidade das inovações obtida na expressão (2.5). Note que,

$$\delta = \mathbb{P}(X_t \neq 0 | X_{t-1} = 0) = \frac{(1 - \pi)\mu}{(1 + \mu)[1 + \alpha(1 + \pi\mu)]},$$

Assim, $\delta = 1 - \mathbb{P}(X_t = 0 | X_{t-1} = 0) = 1 - \left[\frac{\mathbb{P}(X_t=0, X_{t-1}=0)}{\mathbb{P}(X_{t-1}=0)} \right]$. Efetuando algumas operações e isolando α , temos $\alpha = \frac{[\frac{\mu(1-\pi)}{\delta(1+\mu)}]^{-1}}{(1+\pi\mu)} = \frac{\mu(1-\pi) - \delta(1+\mu)}{\delta(1+\mu)(1+\pi\mu)}$, implica que o estimador de probabilidade do parâmetro α é dado por

$$\hat{\alpha} = \frac{\hat{\mu}_{PB}(1 - \hat{\pi}_{PB}) - \hat{\delta}(1 + \hat{\mu}_{PB})}{\hat{\delta}(1 + \hat{\mu}_{PB})(1 + \hat{\mu}_{PB}\hat{\pi}_{PB})},$$

em que

$$\hat{\delta} = 1 - \frac{\sum_{t=2}^n \mathbb{I}_{\{X_t=0, X_{t-1}=0\}}}{\sum_{t=1}^n \mathbb{I}_{\{X_t=0\}}}.$$

Proposição 2.2.2.4.1. *Os estimadores $\hat{\alpha}_{PB}$, $\hat{\mu}_{PB}$ e $\hat{\pi}_{PB}$ são consistentes para estimar α , μ e π , respectivamente.*

Demonstração. O Processo ZMGINAR(1) é um processo Markoviano, Estacionário e Ergódico. Dessa forma, as estatísticas $n^{-1} \sum_{t=1}^n \mathbb{I}_{\{X_t > 0\}}$, $n^{-1} \sum_{t=1}^n \mathbb{I}_{\{X_t > 1\}}$, $n^{-1} \sum_{t=1}^n \mathbb{I}_{\{X_t = 1\}}$ e $\hat{\rho}(1)$ são estimadores consistentes para $\mathbb{P}(X_t > 0)$, $\mathbb{P}(X_t > 1)$, $\mathbb{P}(X_t = 1)$ e $\rho(1)$, respectivamente. \square

2.2.3 Estudo de Simulação de Monte Carlo

Nesta subseção, serão apresentadas os resultados das simulações referente à estimação dos parâmetros do modelo ZMGINAR(1). Os resultados da simulação são expostos por meio de tabela e gráficos. Através de 5000 replicações de Monte Carlo, avaliamos o desempenho dos estimadores em tamanho amostral finito e sob diferentes cenários. Os tamanhos amostrais considerados foram $n = 100, 200, 400$ e 800 . Utilizando as mesmas configurações de Barreto-Souza (2015) para o modelo ZMGINAR(1): (a) $\alpha = 0.5, \mu = 2$ e $\pi = -0.4$, (b) $\alpha = 0.5, \mu = 2$ e $\pi = -0.2$, (c) $\alpha = 0.5, \mu = 2$ e $\pi = 0.1$, (d) $\alpha = 0.5, \mu = 2$ e $\pi = 0.2$ e (e) $\alpha = 0.5, \mu = 2$ e $\pi = 0.3$.

As médias das estimativas e os erros quadráticos médios (EQM), entre parênteses, para avaliar a qualidade das estimativas para cada combinação dos parâmetros α, μ e π estão presentes na Tabela 2.1.

Pode-se notar, de forma geral, que todos os estimadores atingiram bons resultados em termos de viés e EQM, próximos dos parâmetros verdadeiros. Contudo, alguns estimadores apresentaram EQM grandes. Por exemplo, no cenário (a), para $n = 100$, a estimativa do parâmetro α pelo método de estimação baseada em Probabilidade, é dada por

0.473 e EQM igual a 1.629, ou seja, este estimador apresenta uma maior variabilidade para o parâmetro α em comparação aos demais estimadores, sendo notório na Figura 2.4. Observamos também que à medida que o tamanho da amostra aumenta, o EQM dos quatro estimadores diminui tendendo para zero, evidenciando que os quatro estimadores são consistentes. Na Tabela 2.1 observa-se, de modo geral, que o estimador de máxima verossimilhança obteve melhores resultados, tanto em termos de viés, quanto no EQM, por exemplo, no caso (b), para $n = 200$, as estimativas para os parâmetros α , μ e π e EQM pelos métodos de MV, YW, PB e MQC são, respectivamente, 0.491 (0.005), 0.479 (0.006), 0.490 (0.054) e 0.481 (0.006) para o parâmetro α , -0.209 (0.005), -0.236 (0.019), -0.206 (0.008) e -0.236 (0.021) para o parâmetro π e pelos métodos de MV, YW e MQ para o parâmetro μ são, respectivamente, 1.989 (0.093), 1.965 (0.119) e 2.056 (0.224), ou seja, nota-se que a estimativa pelo método de MV, está mais próxima do verdadeiro parâmetro e, o EQM mais próximo de zero, evidenciando a eficácia do estimador proposto em relação aos demais. Portanto, para os parâmetros α e μ , o estimador de MV obteve melhores resultados em termos de viés e EQM em todas as situações e para o parâmetro π tanto o estimador de MV, quanto o PB, obteve bons resultados em termos de EQM e viés. Para confirmar o desempenho dos estimadores são realizados gráficos mostrados na Figura 2.4 que foram obtidos com o tamanho amostral $n = 400$ para todos os cenários. Note que, o EQM para o estimador de MV é menor que os dos outros estimadores e para o parâmetro π , percebe-se que tanto o MV, quanto o PB estão bem próximos das estimativas, mas o MV tem um EQM menor. Portanto, recomenda-se o uso do estimador de máxima verossimilhança para os parâmetros α , μ e π , respectivamente, tendo um ótimo desempenho em termos de EQM e viés.

Tabela 2.1: Estimativas médias e EQM (entre parênteses) provenientes dos resultados numéricos da simulação para valores verdadeiros de α, μ e π .

n	Estimador de α				Estimador de μ			Estimador de π			
	$\hat{\alpha}_{MV}$	$\hat{\alpha}_{YW}$	$\hat{\alpha}_{PB}$	$\hat{\alpha}_{MQC}$	$\hat{\mu}_{MV}$	$\hat{\mu}_{YW}$	$\hat{\mu}_{PB}$	$\hat{\pi}_{MV}$	$\hat{\pi}_{YW}$	$\hat{\pi}_{PB}$	$\hat{\pi}_{MQC}$
a) $\alpha = 0.5, \mu = 2 \text{ e } \pi = -0.4$											
100	0.494 (0.013)	0.459 (0.013)	0.473 (1.629)	0.465 (0.013)	2.010 (0.210)	1.948 (0.214)	2.141 (0.544)	-0.420 (0.014)	-0.476 (0.044)	-0.415 (0.021)	-0.476 (0.050)
200	0.492 (0.005)	0.478 (0.007)	0.486 (0.492)	0.480 (0.007)	1.989 (0.087)	1.965 (0.110)	2.056 (0.221)	-0.413 (0.005)	-0.440 (0.021)	-0.409 (0.010)	-0.440 (0.024)
400	0.496 (0.002)	0.490 (0.003)	0.485 (0.218)	0.491 (0.003)	1.996 (0.043)	1.985 (0.056)	2.027 (0.101)	-0.406 (0.002)	-0.420 (0.010)	-0.405 (0.005)	-0.420 (0.011)
800	0.498 (0.001)	0.494 (0.002)	0.501 (0.095)	0.495 (0.002)	1.996 (0.022)	1.990 (0.028)	2.011 (0.051)	-0.403 (0.001)	-0.411 (0.005)	-0.403 (0.002)	-0.411 (0.005)
b) $\alpha = 0.5, \mu = 2 \text{ e } \pi = -0.2$											
100	0.488 (0.014)	0.459 (0.013)	0.483 (0.123)	0.463 (0.013)	1.988 (0.206)	1.924 (0.228)	2.108 (0.509)	-0.220 (0.010)	-0.274 (0.040)	-0.216 (0.016)	-0.273 (0.046)
200	0.491 (0.005)	0.479 (0.006)	0.489 (0.054)	0.481 (0.006)	1.997 (0.093)	1.970 (0.119)	2.065 (0.224)	-0.209 (0.005)	-0.236 (0.019)	-0.206 (0.008)	-0.236 (0.021)
400	0.495 (0.002)	0.488 (0.003)	0.499 (0.027)	0.489 (0.003)	1.994 (0.045)	1.981 (0.060)	2.028 (0.107)	-0.205 (0.002)	-0.219 (0.009)	-0.204 (0.004)	-0.220 (0.010)
800	0.498 (0.001)	0.495 (0.002)	0.499 (0.014)	0.496 (0.002)	1.996 (0.023)	1.990 (0.031)	2.012 (0.049)	-0.202 (0.001)	-0.208 (0.005)	-0.201 (0.002)	-0.208 (0.005)
c) $\alpha = 0.5, \mu = 2 \text{ e } \pi = 0.1$											
100	0.486 (0.014)	0.460 (0.014)	0.482 (0.047)	0.465 (0.013)	2.008 (0.287)	1.926 (0.277)	2.134 (0.633)	0.091 (0.011)	0.039 (0.033)	0.089 (0.016)	0.039 (0.038)
200	0.490 (0.005)	0.478 (0.007)	0.490 (0.023)	0.481 (0.007)	1.995 (0.103)	1.958 (0.145)	2.073 (0.264)	0.094 (0.005)	0.065 (0.017)	0.094 (0.008)	0.065 (0.018)
400	0.495 (0.002)	0.489 (0.003)	0.497 (0.011)	0.491 (0.003)	1.989 (0.051)	1.969 (0.074)	2.021 (0.118)	0.098 (0.002)	0.082 (0.008)	0.097 (0.004)	0.082 (0.009)
800	0.498 (0.001)	0.496 (0.002)	0.497 (0.005)	0.496 (0.002)	2.001 (0.026)	1.994 (0.038)	2.017 (0.059)	0.099 (0.001)	0.093 (0.004)	0.099 (0.002)	0.093 (0.004)
d) $\alpha = 0.5, \mu = 2 \text{ e } \pi = 0.2$											
100	0.487 (0.015)	0.460 (0.014)	0.475 (0.045)	0.465 (0.013)	2.022 (0.330)	1.914 (0.290)	2.172 (0.738)	0.191 (0.011)	0.133 (0.033)	0.187 (0.017)	0.132 (0.038)
200	0.490 (0.005)	0.479 (0.007)	0.487 (0.021)	0.481 (0.007)	2.002 (0.106)	1.962 (0.158)	2.075 (0.285)	0.196 (0.005)	0.168 (0.016)	0.194 (0.008)	0.169 (0.017)
400	0.496 (0.003)	0.491 (0.004)	0.492 (0.010)	0.492 (0.004)	1.999 (0.056)	1.982 (0.089)	2.037 (0.130)	0.199 (0.003)	0.185 (0.008)	0.199 (0.004)	0.185 (0.008)
800	0.498 (0.001)	0.495 (0.002)	0.497 (0.005)	0.495 (0.002)	2.001 (0.027)	1.989 (0.041)	2.022 (0.064)	0.199 (0.001)	0.190 (0.004)	0.199 (0.002)	0.190 (0.004)
e) $\alpha = 0.5, \mu = 2 \text{ e } \pi = 0.3$											
100	0.486 (0.017)	0.457 (0.015)	0.469 (0.042)	0.461 (0.015)	2.014 (0.388)	1.899 (0.323)	2.152 (0.752)	0.295 (0.011)	0.241 (0.030)	0.289 (0.017)	0.241 (0.034)
200	0.488 (0.005)	0.476 (0.008)	0.483 (0.021)	0.478 (0.007)	1.997 (0.115)	1.947 (0.180)	2.081 (0.323)	0.299 (0.005)	0.268 (0.016)	0.295 (0.008)	0.268 (0.017)
400	0.494 (0.003)	0.488 (0.004)	0.492 (0.010)	0.489 (0.004)	1.996 (0.056)	1.970 (0.091)	2.037 (0.146)	0.299 (0.003)	0.283 (0.007)	0.297 (0.004)	0.283 (0.008)
800	0.497 (0.001)	0.494 (0.002)	0.495 (0.005)	0.495 (0.002)	1.998 (0.029)	1.988 (0.047)	2.017 (0.071)	0.300 (0.000)	0.293 (0.004)	0.300 (0.000)	0.293 (0.004)

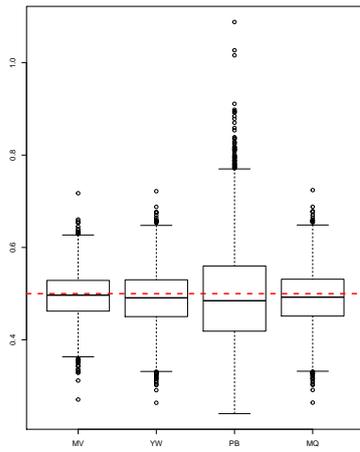
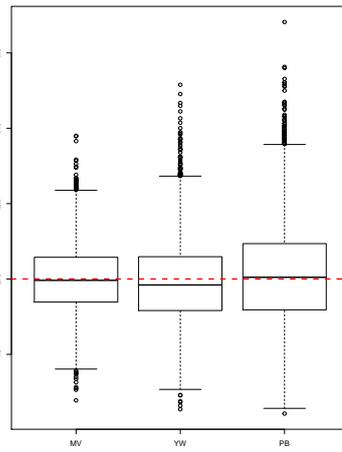
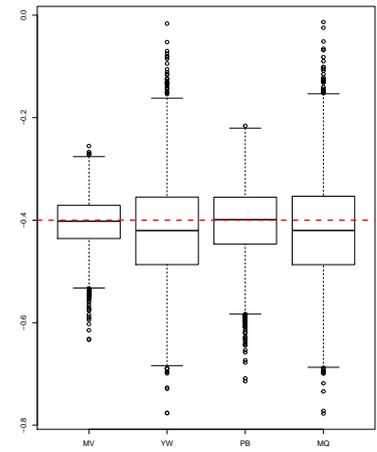
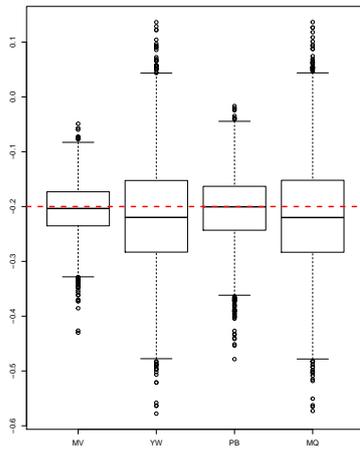
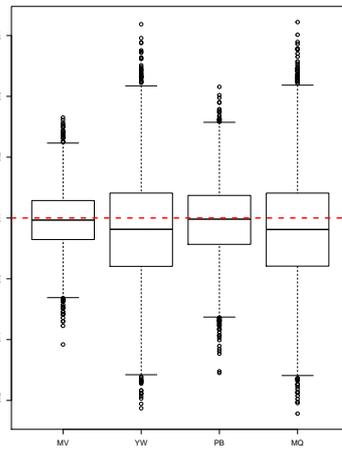
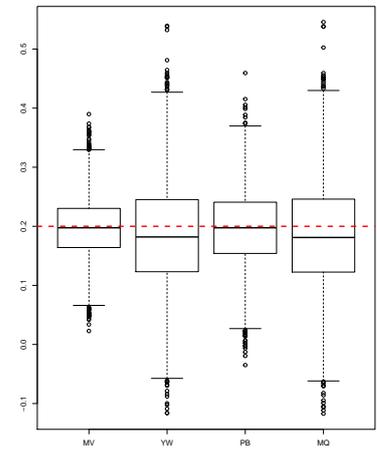
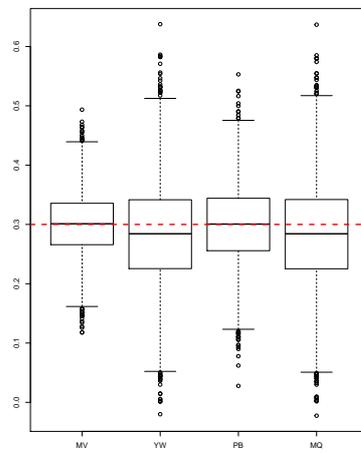
(a) $\alpha = 0.5$ (b) $\mu = 2$ (c) $\pi = -0.4$ (d) $\pi = -0.2$ (e) $\pi = 0.1$ (f) $\pi = 0.2$ (g) $\pi = 0.3$

Figura 2.4: *Box plots* das estimativas dos parâmetros α , μ e π com tamanho da amostra $n = 400$. (a) $\alpha = 0.5$, (b) $\mu = 2$, (c) $\pi = -0.4$, (d) $\pi = -0.2$, (e) $\pi = 0.1$, (f) $\pi = 0.2$ e (g) $\pi = 0.3$.

Capítulo 3

Gráficos de Controle para o processo ZMGINAR(1)

Os gráficos de controle são uma das ferramentas do controle estatístico de qualidade mais conhecidas e difundidas (MONTGOMERY, 2007). Seguindo esta tendência, este capítulo propõe gráficos de controle para monitorar o processo ZMGINAR(1). Concentramos principalmente nos gráficos de controle que podem ser usados para identificar aumento de mudanças na média. Isto não significa que as diminuições na média não sejam relevantes. Como os processos autorregressivos de valores inteiros são frequentemente usados para apresentar o número de defeitos, o número de queixas dos clientes ou o número de pacientes na maioria das aplicações, ou seja, uma diminuição na média geralmente ocorre somente após uma ação corretiva no processo. Existem muitas maneiras pelas quais uma causa atribuível pode tornar este processo fora de controle, por exemplo, segundo Rakitizis, Weiß e Castagliola (2016) a mudança no processo pode-se dar pelos valores esperados de X_t ou ϵ_t , sendo que estes valores podem mudar de um valor em controle (IC) para algum fora de controle (FC), considerando também que as mudanças na correlação de X_t e X_{t-l} , $l \geq 1$, podem ocorrer.

Assumindo que o processo monitorado $\{X_t\}_{t \geq 1}$ segue o modelo ZMGINAR(1), no presente estudo, o objetivo principal é detectar uma mudança no número médio $\mu_X = E(X_t) = \mu(1 - \pi)$ de contagens, e ficamos atentos aos aumentos no parâmetro μ . Por exemplo, se X_t descreve o número de criminosos em uma cidade no momento t , um aumento de μ implicaria em um aumento no número médio de criminosos por algum motivo (LI; WANG; ZHU, 2016). Dessa forma, assumimos que a distribuição marginal

de $\{X_t\}_{t \geq 1}$ tem média μ_X . Quando o processo se encontra em IC (FC), denotamos $\mu_X = \mu_{X_0}$ ($\mu_X = \mu_{X_1}$). Ou seja, os valores em controle ou fora de controle para os parâmetros do modelo ZMGINAR(1) são denotados por μ_0, π_0 e μ_1, π_1 , respectivamente. Uma mudança no parâmetro μ_{X_0} pode ocorrer se um (ou mais) parâmetros do processo mudarem apropriadamente. Mais especificamente, de acordo com a expressão (2.2) para μ_X , sabemos que a média do ZMGINAR(1) aumenta quando o μ aumenta e o π permanece constante, ou quando o π diminui negativamente. Uma abordagem para a detecção de mudanças no aumento em μ_{X_0} é considerar os tradicionais gráficos de Shewhart proposto por Shewhart (1925).

3.1 Gráficos de Shewart

Os gráficos tradicionais de Shewhart propostos por Shewhart (1925) baseiam-se nos princípios de que o limite inferior de controle (LIC) e o limite superior de controle (LSC) são calculados através da média das estatísticas de monitoramento mais ou menos três vezes o desvio padrão (LI; WANG; ZHU, 2016). As realizações do processo $\{X_t\}_{t \geq 1}$ são traçadas no gráfico com um LSC , e quando $X_t > LSC$, o processo está fora de controle.

Uma das maneiras de se avaliar o desempenho de um gráfico de controle é mediante o número médio de amostras (NMA), que corresponde ao número médio de amostras necessárias para se detectar uma mudança ocorrida no processo. É a quantidade de amostras na média, necessárias para que o gráfico de controle sinalize um alarme (FC). Defina L como uma variável aleatória que denota o número de pontos plotados no gráfico até um sinal FC seja sinalizado pela primeira vez. Dessa forma, podemos dizer que $E(L) = NMA$. Assim,

$$NMA_0 = \mathbb{E}(L|\tau = 1) \text{ e } NMA_1 = \lim_{\tau \rightarrow \infty} NMA^{(\tau)},$$

em que $NMA^{(\tau)} = \mathbb{E}(L - \tau + 1|L \geq \tau, \tau)$, $\tau \geq 1$, é o NMA condicional esperado ver (WEIß, 2011). Além disso, $\tau \in \{1, 2, \dots\}$ é ponto de mudança fixo, em que $t < \tau$, o processo é assumido estando em controle, e para $t \geq \tau$, o processo é assumindo como fora de controle. (RAKITZIS; WEIß; CASTAGLIOLA, 2016)

O LSC do gráfico de Shewhart para esse método é na verdade um limite de pro-

babilidade, isto é, procura-se um valor inteiro do LSC de tal forma que o NMA_0 seja o mais próximo possível do desejado, ou seja, não foi calculado usando os limites três vezes o desvio padrão do processo ou o limite simulado. Significa que na matriz de probabilidade de transição, a probabilidade de um ponto estar acima do LSC dado os parâmetros do processo, o NMA_0 dependerá do LSC. Portanto, é comum na prática usar o NMA_0 como uma medida de desempenho quando o processo está sob controle. Em seguida, o número esperado de pontos plotados no gráfico até o alarme falso é avaliado. Quando o processo está fora de controle, o NMA_0 não representa o comportamento fora de controle, porque um processo nem sempre fica fora de controle exatamente no momento em que a monitoração começa. Assim, uma medida de desempenho mais realista é o NMA_1 , que concede uma aproximação do verdadeiro atraso médio para a detecção após o ponto de mudança. Mais informações a respeito desse método veja Weiß e Testik (2009) e Weiß e Testik (2012)

Quando o processo está sob controle denotamos $\mu_X = \mu_{X_0}$ e quando está fora de controle ($\mu_X = \mu_{X_1}$). Assim o NMA para o processo sob IC (FC) são denotados por NMA_0 e (NMA_1). Vale ressaltar que o NMA_0 é o número médio de pontos plotados no gráfico desde o início do monitoramento até que o gráfico detecte um alarme falso, enquanto que, o NMA_1 é o número médio de pontos plotados no gráfico desde o início de um processo até o gráfico detectar o deslocamento fora de controle. A determinação do LSC para o gráfico Shewhart é feita de modo a obter um valor NMA_0 desejado (RAKITZIS; WEIß; CASTAGLIOLA, 2016). Além do gráfico de Shewart, neste estudo, também consideramos os gráficos de controle do tipo CUSUM, proposto pela primeira vez por Page (1954).

3.2 Gráficos de CUSUM

O gráfico de soma acumulada (CUSUM) proposto por Page (1954) é um aprimoramento do gráfico de controle de Shewhart, sendo este, mais apropriado para reconhecer o histórico dos dados característica ausente em gráficos mais simples, e também para reconhecer pequenas alterações nos processos muito antes dos alarmes dos gráficos de Shewhart (SAMOHYL, 2009). Além disso, os gráficos de CUSUM têm propriedades de otimização conhecidas para obter uma mudança sustentada na média para uma distribuição IC para

uma distribuição FC. Mais detalhes, ver Hawkins e Olwell (2012).

Muitas atenções relevantes têm sido voltadas para estes gráficos devido a sua sensibilidade na detecção de pequenas mudanças, em particular, Brook e Evans (1972), Chao-Wen e Jr (2001), Megahed et al. (2011) e Weiß e Testik (2012). O Gráfico de CUSUM foi aplicado para controlar um processo autorregressivo de contagens de Poisson com um número excessivo de zeros por Rakitzis, Weiß e Castagliola (2016). De maneira análoga, neste trabalho aplicamos este método no processo ZMGINAR(1).

O gráfico de CUSUM é definido da seguinte maneira,

$$\begin{aligned} C_0 &= c_0 \\ C_t &= \max(0, C_{t-1} + X_t - k), \quad t \geq 1, \end{aligned} \quad (3.1)$$

em que C_t é o valor plotado no tempo t . O valor de $c_0 \geq 0$ é o valor inicial. Sejam $k, h \in \mathbb{N}$ com $\mu_{X_0} < k < h$. As estatísticas de monitoramento C_t são traçadas no gráfico CUSUM com a região de controle $[0, h]$, isto é, o processo ZMGINAR(1) é considerado estando IC a menos que $C_t \notin [0, h]$. O valor de $h > 0$ é conhecido como o intervalo de decisão, na verdade, é o limite de controle do gráfico CUSUM. Um alarme será sinalizado se a estatística C_t exceder o limite de controle h pela primeira vez. É comum definir o valor inicial de $c_0 = 0$, que pode ser uma resposta inicial mais rápida ao processo quando ele está fora de controle, ver Lucas e Crosier (1982). O *NMA* do gráfico de CUSUM pode ser calculado adaptando a abordagem da cadeia de Markov empregada por Weiß e Testik (2009) e Weiß e Testik (2012). Seguindo esta ideia, a cadeia de Markov é aplicada usando as probabilidades de transição para encontrar o *NMA* de um gráfico. Com efeito, as cadeias de Markov representam um sistema que varia o seu estado ao longo do tempo, tendo as seguintes fases: (Estado 1: Em controle, Estado 2: Fora de controle). A cadeia é constituída por quatro conjuntos de m estados transitórios. O status do processo (em controle ou fora de controle) quando a $(i + 1)$ é-sima amostra é tomada e a posição do ponto de amostra i define os quatro conjuntos de estados transitórios (OVIEDO-TRESPALACIOS et al., 2014).

Os *NMA* do gráfico CUSUM são calculados da seguinte forma. O processo bivariado $(X_1, C_1), (X_2, C_2), \dots$ definidos nas Equações (2.4) e (3.1) é uma cadeia de Markov

bivariada com probabilidades de transição

$$\begin{aligned}
p(a, b|c, d) &= P(X_t = a, C_t = b \mid X_{t-1} = c, C_{t-1} = d) \\
&= \delta_{b, \max(0; a-k+d)} \cdot p_{a|c}, \\
p_1(a, b|c) &= P(X_1 = a, C_1 = b \mid C_0 = c) \\
&= \delta_{b, \max(0; a-k+c)} \cdot p_a,
\end{aligned} \tag{3.2}$$

em que $\delta_{x,y}$ denota o Kronecker delta (cujo valor é 1, se $x = y$ e 0, se $x \neq y$), $p_a = P(X_t = a)$ e $p_{a|c} = P(X_t = a \mid X_{t-1} = c)$. Por causa desta propriedade de Markov, é possível adaptar a abordagem feita por Brook e Evans (1972).

As probabilidades de transição dada em (3.2) são necessários para calcular os NMA do gráfico CUSUM. Porém, necessita-se controlar o conjunto de valores (X_t, C_t) . O tamanho deste conjunto de valores determina a dimensão da matriz \mathbf{Q} para a abordagem de cadeia de Markov. Considere o caso em que $k, h, c_0 \in \mathbb{N}$. Em primeiro instante, todos os pares $\mathbb{N} \times \{0, \dots, h-1\}$ podem ser alcançáveis. Porém, a estatística C_t excede h se $X_t - k + C_{t-1} \geq h$, veja na Equação (3.1). Então o valor de $X_t \geq k + h$ é sempre dado um sinal fora de controle. Os pares (X_t, C_t) são bem restritos, desde de que $C_{t-1} \leq h-1$, apenas o par (n, i) pode ser alcançado da seguinte forma

$$\begin{aligned}
(X_t, C_t) &= (n, i) \\
&\Leftrightarrow X_t = n, C_t = \max(0; n - k + C_{t-1}) = i,
\end{aligned}$$

dessa forma obtém-se a seguinte restrição

$$n \leq i + k \quad \text{e} \quad n \geq i + k - h + 1.$$

Então o par em controle (n, i) é dado por

$$\tau(h, k) = \{(n, i) \mid 0 \leq i \leq h-1, \max(0; i + k - h + 1) \leq n \leq i + k\}, \tag{3.3}$$

com tamanho $|\tau(h, k)| = \frac{1}{2}(h-k)(h+k+1) + hk$.

Sejam $(X_1, C_1), (X_2, C_2), \dots$ um processo bivariado de contagens e a estatística CUSUM definido em (2.4) e (3.1), respectivamente, com $c_0 \in \mathbb{N}$. Seja $h \in \mathbb{N}$ denota o intervalo

de decisão e $k \in \mathbb{N}$ com $\mu_{X_0} \leq k < h$ como valor de referência. Seja $\tau(h, k)$ definido em (3.3) como estado em controle. Seja $\mu_{m,a}$ o número esperado de sinais em controle antes do primeiro sinal fora de controle, dado que $(X_1, C_1) = (m, a) \in \tau(h, k)$. Defina

$$\begin{aligned}\boldsymbol{\mu} &= (\dots, \mu_{n,i}, \dots)^\top, \\ \mathbf{Q}^\top &= (p(n, i \mid m, a))_{(n,i),(m,a) \in \tau(h,k)}.\end{aligned}$$

As dimensões de \mathbf{Q} e $\boldsymbol{\mu}$ são determinadas na Equação (3.3) e as probabilidades de transição dadas em (3.2) que foram calculadas em (2.6), então a solução da equação linear $(\mathbf{I} - \mathbf{Q}) \cdot \boldsymbol{\mu} = \mathbf{1}$, com $\mathbf{I} = (1, 1, \dots, 1)^\top$. O NMA dependendo da escolha do $C_0 = c_0$, é definido como

$$NMA(c_0) = 1 + \sum_{(m,a) \in \tau(h,k)} \mu_{m,a} \cdot p_1(m, a \mid c_0),$$

em que $p_1(m, a \mid c_0)$ é calculado a partir da Equação (3.2).

A análise dos gráficos citadas requer a determinação dos parâmetros de concepção correspondentes (isto é, LSC para os gráficos de Shewhart e (k, h) para o gráfico CUSUM) de modo a ter o valor NMA_0 desejado. Como estamos lidando com séries temporais de valores inteiros, nem sempre é possível obter NMA_0 desejado, isto acontece em função do processo ser discreto o que faz com que no NMA_0 variem e não fiquem tão próximos. Devido a isso, escolhe-se o valor do LSC de modo a obter um NMA_0 mais próximo possível do desejado. Para o gráfico CUSUM, o par de parâmetros (k, h) devem ser determinados, dados os respectivos valores do parâmetro do IC, isto é, μ_0 e π_0 . De acordo com Hawkins e Olwell (2012), o valor de k pode ser especificado de modo a aumentar a sensibilidade do gráfico CUSUM na detecção da mudança. Portanto, o desempenho para este gráfico pode ser obtido procurando um par ideal (k^*, h^*) , tal que, o NMA_0 em IC seja o mais próximo possível do desejado e para um valor FC específico $\mu_X = \mu_{X_1}$, o par (k^*, h^*) proporciona o valor mínimo do NMA quando está FC. Mencionamos que, para o processo ZMGINAR(1) comporta apenas valores inteiros, procede-se apenas em valores inteiros para o LSC de Shewhart e os valores (k, h) do gráfico CUSUM. Em consequência, o número de estados no cálculo da cadeia de Markov é limitado e os cálculos tornam-se mais simples.

3.3 Estudo Numérico

Nesta seção, apresentaremos os resultados do estudo numérico sobre o desempenho dos gráficos de controle de Shewhart e CUSUM quando as observações vêm do processo ZMGINAR(1). Como já foi mencionado, estamos interessados em detectar mudanças no parâmetro μ_X da distribuição marginal de X_t . Com isso, assumimos que a mudança não ocorra no parâmetro π e permaneça inalterada até ser detectada um sinal de fora controle. As mudanças consideradas em μ_{X_0} são, $\mu_X = \mu_{X_0} + \delta$, com $\delta = 0.1, 0.2, 0.3, 0.4, 0.5, 1.0, 1.5, 2.0$ e consideramos $\alpha = 0.3$ e $\alpha = 0.5$. Os parâmetros para as configurações apresentadas foram determinados de forma a obter o mesmo desempenho em controle entre eles.

Com respeito ao gráfico de CUSUM, Weiß e Testik (2009) forneceram uma regra para escolher o valor de referência k , isto é, $k = \lfloor \mu_{X_0} \rfloor + 1$ ou $k = \lfloor \mu_{X_0} \rfloor + 2$ a função piso em que é o maior número inteiro menor ou igual a este número, por exemplo, o piso de $3.99 = 3$. Em seguida, o valor de h é determinado com um par (k, h) apropriado que dá um valor NMA_0 o mais próximo possível do valor NMA_0 das outras configurações em comparação. Utilizando o método de cadeia de Markov a escolha do h avalia-se o desempenho do NMA no gráfico de CUSUM para vários valores de h , tal que, $h \in \{k, k+1, k+2, \dots\}$ lembrando que k é um número inteiro. Assim, temos pares de valores (k, h) como o par de valores que lhe forneceu esta configuração, ou seja, com o valor do NMA mais próximo para NMA_0 desejado. O menor valor fora de controle do NMA é indicado com texto em negrito. Para todos os cálculos, foi utilizado o software R (R Core Team, 2013).

Para as Tabelas a seguir, apresentamos os perfis dos NMA dos gráficos de controle de Shewhart e CUSUM (colunas denominadas “*c-chart*” e “CUSUM”, respectivamente, para o processo ZMGINAR(1)). Para cada combinação dos parâmetros do processo, diferentemente do que foi realizado em Rakitzis, Weiß e Castagliola (2016) consideramos além do cenário de inflação de zeros (π positivo) e o cenário com deflação de zeros (π negativo), também fornecemos a porcentagem de zeros (p_0), que é calculado a partir da Equação (2.1) já que conhecemos a distribuição marginal do processo, sem a necessidade de simular a porcentagem de zeros como foi realizado em Rakitzis, Weiß e Castagliola (2016). Os valores dos parâmetros para cada gráfico de controle são dados nas linhas intituladas *LSC* (para o gráfico de Shewhart) e h e k (para o gráfico de CUSUM). Para $\delta = 0$, o

valor do NMA_0 é dado para que todos tenham (aproximadamente) o mesmo desempenho do gráfico em controle. Para o desempenho dos gráficos são fornecidos quatro Tabelas, dentre elas, as Tabelas 3.1 e 3.2 apresentam os perfis dos NMA dos gráficos de Shewhart e CUSUM na detecção da mudança em μ_{X_0} para vários valores do parâmetro π submetidos a dados com subdispersão, e nas Tabelas 3.3 e 3.4 diz respeito à mudança na média do processo para vários valores do parâmetro π submetidos a dados com sobredispersão e equidispersão. Os resultados das Tabelas 3.1 e 3.2 mostram, para o caso $\alpha = 0.3$ e $\pi = 0.05, 0.1$ ou 0.2 , que o gráfico de Shewhart é o melhor gráfico na detecção de mudanças na média μ_{X_0} . Nota-se também, quando a proporção de zeros (p_0) é alta, ou seja, aproximadamente, acima de 40%, o gráfico de controle de Shewhart é melhor comparado ao do CUSUM. No entanto, quando tratamos o p_0 inferior a 40%, o gráfico de CUSUM torna-se mais eficiente, por exemplo, $\mu_{X_0} = 2$, $\pi = 0.05$, resultando em um $p_0 = 36.66\%$. Isso mostra de fato que o gráfico CUSUM baseia-se na soma acumulada, e quando se trata de uma maior proporção de zeros (p_0) o gráfico é ineficiente para detectar uma mudança em comparação com o gráfico de Shewhart. Conseqüentemente, quando $\alpha = 0.5$ na Tabela 3.2, o gráfico de Shewhart continua com o melhor desempenho com o p_0 maior que 40%. No entanto, quando o $p_0 = 36.66\%$, o gráfico CUSUM é o melhor gráfico na detecção de mudanças pequenas e moderadas ($\delta < 1.0$) na média do processo (μ_{X_0}). Já nas Tabelas 3.3 e 3.4, os resultados mostram o cenário com deflação de zeros. Verifica-se que a superioridade dos gráficos de Shewhart nas Tabelas 3.1 e 3.2 já não é tão evidente, isso porque a porcentagem de zeros não é tão grande. No entanto, o gráfico de Shewhart continua com o melhor desempenho para mudanças na média de $\mu_{X_0} = 0.5$ e $\mu_{X_0} = 1$ para valores de $\pi = -0.05, -0.1$ e alguns cenários com $\pi = -0.2$ e diferente dos cenários com π positivo, com o π negativo, os NMA_0 dos gráficos são altos, ou seja, aumentando a taxa de alarmes falsos, além disso, mesmo com a porcentagem de zeros sendo grande em alguns cenários, o gráfico de Shewhart não é tão eficiente. Por exemplo, na Tabela 3.3, para $\mu_{X_0} = 0.5$ e $\pi = -0.3$, mesmo a com porcentagem de zeros de $p_0 = 56\%$, o gráfico CUSUM mostrou-se eficiente e com um bom desempenho para deslocamentos ($\delta \geq 1.0$), enquanto que o de Shewhart tornou-se eficiente para deslocamentos ($\delta \leq 0.5$). O gráfico de CUSUM, mostrou-se eficiente mesmo com a porcentagem de zeros alta em relação aos cenários da Tabela 3.1 e 3.2, por exemplo, para $\mu_{X_0} = 1$ e $\pi = -0.2$ resultando em $p_0 = 40\%$, o gráfico de CUSUM, mostrou-se um bom desempenho em comparação ao

de Shewhart na presença de deflação de zeros. Conseqüentemente, na Tabela 3.4 com o cenário com $\alpha = 0.5$ os gráficos de Shewhart e CUSUM alternaram bons desempenhos em relação aos cenários apresentados. Por exemplo, para $\mu_{X_0} = 0.5$ e $\pi = -0.05$ resultando em $p_0 = 65\%$, o gráfico de CUSUM mostrou-se eficiente com deslocamento de $\delta \leq 1.0$ em relação ao de Shewhart mesmo com a porcentagem alta de zeros. Isso evidencia que as diferenças dos cenários com inflação e deflação de zeros, o CUSUM mostrou-se mais eficiente no cenário com deflação de zeros e o de Shewhart com inflação de zeros em determinados casos.

Tabela 3.1: *NMA* dos gráficos de Shewhart e CUSUM do processo ZMGINAR(1), $\alpha = 0.3$.

δ	$\mu_{X_0} = 0.5$		$\mu_{X_0} = 1$		$\mu_{X_0} = 2$	
	$\pi = 0.05, p_0 = 68.33\%$	CUSUM	$\pi = 0.05, p_0 = 52.5\%$	CUSUM	$\pi = 0.05, p_0 = 36.66\%$	CUSUM
0.0	279.129	326.365	173.655	173.416	317.567	310.967
0.1	156.851	181.52	120.702	128.612	254.792	236.355
0.2	99.462	113.542	93.245	97.852	207.974	184.381
0.3	68.699	77.371	73.871	76.415	172.390	147.214
0.4	50.530	56.229	59.874	61.168	144.868	119.998
0.5	38.983	42.943	49.524	50.081	123.244	99.633
1.0	16.402	17.574	24.105	23.867	63.547	48.658
1.5	10.012	10.642	14.712	14.962	38.987	30.142
2.0	7.262	7.707	10.747	10.877	26.802	21.447
<i>LSC</i>	5		7		14	
<i>h</i>		4		10		22
<i>k</i>		2		2		3
δ	$\pi = 0.1, p_0 = 70\%$		$\pi = 0.1, p_0 = 55\%$		$\pi = 0.1, p_0 = 40\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	418.248	412.109	221.885	224.694	290.141	295.365
0.1	190.276	307.601	139.853	181.616	232.889	226.13
0.2	130.283	222.918	111.532	145.95	190.280	177.516
0.3	93.115	159.600	90.086	117.522	157.900	142.501
0.4	69.348	115.460	73.820	95.371	132.857	116.694
0.5	53.572	85.694	61.387	78.278	113.180	97.273
1.0	21.700	30.234	29.644	35.506	58.815	48.094
1.5	12.733	17.248	18.084	21.021	36.394	29.96
2.0	8.970	12.193	12.717	14.655	25.237	21.387
<i>LSC</i>	5		7		14	
<i>h</i>		10		9		20
<i>k</i>		1		2		3

δ	$\mu_{X_0} = 0.5$ $\pi = 0.2, p_0 = 73.33\%$		$\mu_{X_0} = 1$ $\pi = 0.2, p_0 = 60\%$		$\mu_{X_0} = 2$ $\pi = 0.2, p_0 = 46.66\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	329.906	348.133	184.878	202.387	375.831	377.652
0.1	185.135	193.577	133.678	144.753	301.410	291.936
0.2	117.249	120.727	100.821	108.277	245.926	230.962
0.3	80.886	82.181	78.795	84.016	203.763	186.538
0.4	59.424	59.843	63.409	67.199	171.159	153.465
0.5	45.794	45.92	52.286	55.129	145.547	128.354
1.0	19.181	19.611	25.698	26.706	74.881	63.715
1.5	11.673	12.379	16.328	16.901	45.840	39.441
2.0	8.450	9.238	11.929	12.35	31.448	27.914
<i>LSC</i>	5		7		14	
<i>h</i>		7		5		20
<i>k</i>		1		3		3

Fonte: Elaborada pelo autor

Tabela 3.2: *NMA* dos gráficos de Shewhart e CUSUM do processo ZMGINAR(1), $\alpha = 0.5$.

δ	$\mu_{X_0} = 0.5$ $\pi = 0.05, p_0 = 68.33\%$		$\mu_{X_0} = 1$ $\pi = 0.05, p_0 = 52.5\%$		$\mu_{X_0} = 2$ $\pi = 0.05, p_0 = 36.66\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	120.777	130.561	176.374	176.99	168.064	172.417
0.1	77.402	85.665	129.656	129.047	140.350	139.746
0.2	54.326	60.907	99.158	97.846	118.733	115.36
0.3	40.767	45.992	78.442	76.686	101.757	96.814
0.4	32.129	36.369	63.799	61.817	88.221	82.464
0.5	26.277	29.814	53.101	51.041	77.278	71.185
1.0	13.426	15.39	26.954	25.334	45.015	39.918
1.5	9.141	10.582	17.416	16.34	30.280	26.793
2.0	7.112	8.289	12.829	12.118	22.358	20.051
<i>LSC</i>	4		7		12	
<i>h</i>		6		11		14
<i>k</i>		1		2		3
δ	$\pi = 0.1, p_0 = 70\%$		$\pi = 0.1, p_0 = 55\%$		$\pi = 0.1, p_0 = 40\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	127.21	131.294	185.577	200.031	379.039	393.876
0.1	81.462	84.646	136.347	146.136	305.759	305.586
0.2	57.148	59.39	104.234	110.973	250.802	242.324

Tabela 3.2 – Continuação

δ	$\mu_{X_0} = 0.5$ $\pi = 0.1, p_0 = 70\%$		$\mu_{X_0} = 1$ $\pi = 0.1, p_0 = 55\%$		$\mu_{X_0} = 2$ $\pi = 0.1, p_0 = 40\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.3	42.866	44.375	82.426	87.074	208.895	196.03
0.4	33.770	34.787	67.015	70.248	176.376	161.483
0.5	27.610	28.309	55.759	58.035	150.745	135.229
1.0	14.092	14.277	28.262	28.821	79.413	67.72
1.5	9.591	9.72	18.242	18.578	49.633	42.454
2.0	7.462	7.595	13.426	13.774	34.664	30.441
<i>LSC</i>	3		7		14	
<i>h</i>		4		7		34
<i>k</i>		1		2		3

δ	$\mu_{X_0} = 0.5$ $\pi = 0.2, p_0 = 73.33\%$		$\mu_{X_0} = 1$ $\pi = 0.2, p_0 = 60\%$		$\mu_{X_0} = 2$ $\pi = 0.2, p_0 = 46.66\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	403.459	382.018	207.396	215.053	197.311	181.866
0.1	228.508	223.508	152.208	159.565	164.584	151.599
0.2	145.857	145.161	116.265	122.833	139.123	128.279
0.3	101.349	101.782	91.868	97.531	119.139	110.017
0.4	74.949	75.631	74.638	79.497	103.211	95.504
0.5	58.106	58.789	62.057	66.262	90.341	83.814
1.0	24.906	25.36	31.363	33.876	52.441	49.661
1.5	15.387	15.808	20.202	22.161	35.170	34.258
2.0	11.257	11.699	14.849	16.574	25.903	26.014
<i>LSC</i>	5		7		12	
<i>h</i>		4		10		21
<i>k</i>		2		2		3

Fonte: Elaborada pelo autor

Tabela 3.3: *NMA* dos gráficos de Shewhart e CUSUM do processo ZMGINAR(1), $\alpha = 0.3$.

δ	$\mu_{X_0} = 0.5$ $\pi = -0.05, p_0 = 65\%$		$\mu_{X_0} = 1$ $\pi = -0.05, p_0 = 47.5\%$		$\mu_{X_0} = 2$ $\pi = -0.05, p_0 = 30\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	253.309	288.207	131.997	131.784	193.674	204.232
0.1	142.465	160.312	96.247	93.84	158.047	157.368
0.2	90.411	100.332	73.149	70.026	131.029	124.407
0.3	62.497	68.428	57.530	54.314	110.190	100.615
0.4	46.002	49.782	46.536	43.499	93.851	83.033
0.5	35.515	38.063	38.536	35.783	80.850	69.759
1.0	14.987	15.672	19.161	17.766	43.907	35.814

Tabela 3.3 – Continuação

δ	$\mu_{X_0} = 0.5$ $\pi = -0.05, p_0 = 65\%$		$\mu_{X_0} = 1$ $\pi = -0.05, p_0 = 47.5\%$		$\mu_{X_0} = 2$ $\pi = -0.05, p_0 = 30\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
1.5	9.167	9.539	12.202	11.551	28.000	23.001
2.0	6.660	6.934	8.894	8.622	19.832	16.793
<i>LSC</i>	5		7		13	
<i>h</i>		4		8		22
<i>k</i>		2		2		3
δ	$\mu_{X_0} = 0.5$ $\pi = -0.1, p_0 = 63.3\%$		$\mu_{X_0} = 1$ $\pi = -0.1, p_0 = 45\%$		$\mu_{X_0} = 2$ $\pi = -0.1, p_0 = 26.66\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	242.151	274.525	246.877	268.287	275.102	284.251
0.1	136.247	142.792	171.554	175.919	220.806	201.963
0.2	86.499	85.181	125.091	122.324	180.299	149.08
0.3	59.815	56.461	94.868	89.348	149.506	113.862
0.4	44.044	40.554	74.314	68.016	125.685	89.628
0.5	34.015	30.963	59.806	53.606	106.966	72.445
1.0	14.375	13.522	26.763	23.312	55.263	33.423
1.5	8.802	8.785	15.937	14.263	33.971	20.683
2.0	6.399	6.685	11.094	10.309	23.396	14.868
<i>LSC</i>	4		8		14	
<i>h</i>		8		11		28
<i>k</i>		1		2		3
δ	$\pi = -0.2, p_0 = 60\%$		$\pi = -0.2, p_0 = 40\%$		$\pi = -0.2, p_0 = 20\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	222.610	243.46	226.890	221.118	170.059	166.827
0.1	125.355	157.739	157.739	144.91	138.837	126.587
0.2	79.644	115.068	115.068	100.846	115.146	98.952
0.3	55.115	87.305	87.305	73.804	96.870	79.408
0.4	40.613	68.420	68.420	56.34	82.538	65.213
0.5	31.386	55.087	55.087	44.555	71.132	54.65
1.0	13.302	24.702	24.702	19.756	38.705	28.277
1.5	8.161	14.736	14.736	12.288	24.728	18.51
2.0	5.942	10.272	10.272	8.989	17.544	13.758
<i>LSC</i>	5		8		12	
<i>h</i>		4		11		23
<i>k</i>		2		2		3
δ	$\pi = -0.3, p_0 = 56.66\%$		$\pi = -0.3, p_0 = 35\%$		$\pi = -0.3, p_0 = 13.33\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	206.057	219.685	209.950	210.485	233.618	244.144
0.1	116.125	122.201	146.026	143.035	187.597	171.299

Tabela 3.3 – Continuação

δ	$\mu_{X_0} = 0.5$		$\mu_{X_0} = 1$		$\mu_{X_0} = 2$	
	$\pi = -0.3, p_0 = 56.66\%$		$\pi = -0.3, p_0 = 35\%$		$\pi = -0.3, p_0 = 13.33\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.2	73.833	76.563	106.567	102.326	153.247	125.763
0.3	51.130	52.311	80.889	76.38	127.130	96.086
0.4	37.702	38.143	63.418	59.069	106.922	75.988
0.5	29.155	29.238	51.080	47.063	91.039	61.894
1.0	12.389	12.202	22.950	20.654	47.143	30.128
1.5	7.616	7.513	13.713	12.395	29.045	19.547
2.0	5.553	5.512	9.572	8.77	20.046	14.547
<i>LSC</i>	5		8		14	
<i>h</i>		4		7		31
<i>k</i>		2		3		3

Fonte: Elaborada pelo autor

Tabela 3.4: *NMA* dos gráficos de Shewhart e CUSUM do processo ZMGINAR(1), $\alpha = 0.5$.

δ	$\mu_{X_0} = 0.5$		$\mu_{X_0} = 1$		$\mu_{X_0} = 2$	
	$\pi = -0.05, p_0 = 65\%$		$\pi = -0.05, p_0 = 47.5\%$		$\pi = -0.05, p_0 = 30\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	310.721	338.123	307.654	309.319	327.655	337.719
0.1	176.551	199.493	215.555	221.034	264.522	263.025
0.2	112.999	130.096	158.300	164.54	217.125	209.435
0.3	78.718	91.361	120.873	126.761	180.968	170.158
0.4	58.352	67.89	95.299	100.532	152.900	140.794
0.5	45.337	52.725	77.164	81.726	130.769	118.435
1.0	19.595	22.517	35.430	37.863	69.114	60.585
1.5	12.165	13.849	21.502	23.113	43.325	38.63
2.0	8.924	10.099	15.178	16.415	30.338	28.034
<i>LSC</i>	4		8		14	
<i>h</i>		5		9		34
<i>k</i>		2		3		3
δ	$\pi = -0.1, p_0 = 63.3\%$		$\pi = -0.1, p_0 = 45\%$		$\pi = -0.1, p_0 = 26.66\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	104.959	106.598	294.447	292.781	214.134	206.796
0.1	67.426	69.693	206.389	205.87	40.493	38.283
0.2	47.395	49.472	151.622	151.022	35.261	33.279
0.3	35.613	37.351	115.815	114.925	31.038	29.266
0.4	28.100	29.561	91.342	90.276	27.585	26.005
0.5	23.006	24.269	73.984	72.884	24.726	23.32

Tabela 3.4 – Continuação

δ	$\mu_{X_0} = 0.5$ $\pi = -0.1, p_0 = 63.33\%$		$\mu_{X_0} = 1$ $\pi = -0.1, p_0 = 45\%$		$\mu_{X_0} = 2$ $\pi = -0.1, p_0 = 26.66\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
1.0	11.800	12.653	34.020	33.636	15.820	15.062
1.5	8.052	8.773	20.671	20.958	11.389	11.009
2.0	6.272	6.911	14.605	15.251	8.844	8.687
<i>LSC</i>	4		8		13	
<i>h</i>		6		15		29
<i>k</i>		1		2		3
δ	$\pi = -0.2, p_0 = 60\%$		$\pi = -0.2, p_0 = 40\%$		$\pi = -0.2, p_0 = 20\%$	
	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM	<i>c-chart</i>	CUSUM
0.0	273.538	288.828	271.291	258.848	288.895	291.677
0.1	155.712	170.253	190.313	184.9	233.400	222.187
0.2	99.817	111.002	139.907	137.644	191.695	173.826
0.3	69.639	77.973	106.937	106.076	159.869	139.33
0.4	51.693	57.976	84.395	84.177	135.155	114.147
0.5	40.216	45.064	68.400	68.485	115.661	95.363
1.0	17.473	19.351	31.541	31.906	61.305	48.545
1.5	10.888	11.965	19.208	19.598	38.531	31.42
2.0	8.008	8.763	13.596	13.997	27.044	23.189
<i>LSC</i>	5		7		14	
<i>h</i>		5		9		38
<i>k</i>		2		3		3

Fonte: Elaborada pelo autor

Capítulo 4

Aplicações

Para ilustrar a aplicação dos gráficos de controle de Shewhart e CUSUM no monitoramento do modelo ZMGINAR(1), será aplicado a dois conjuntos de dados, um com inflação de zeros e outro com deflação de zeros. O modelo ZMGINAR(1) será comparado com modelo NGINAR(1) (caso particular). As estimativas dos parâmetros dos modelos ZMGINAR(1) e NGINAR(1) foram obtidas através do método de máxima verossimilhança.

4.1 Inflação de Zeros

O primeiro banco de dados analisado representa o número de pessoas que bebem em local público reportados pelo 51º departamento de polícia de Pittsburgh - Estados Unidos, no período de Janeiro de 1990 à Dezembro 2001, disponível <http://www.forecastingprinciples.com> totalizando 144 observações. A Figura 4.1 mostra o gráfico da série, a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF). Nota-se que a utilização de modelos autorregressivos de primeira ordem é adequada, pois no gráfico da PACF apenas a primeira defasagem se destaca dos demais, e no gráfico da ACF nota-se um decaimento de forma exponencial. Na Tabela 4.1 apresenta-se as estatísticas descritivas, e nela observa-se que a variância é um pouco maior que a média, e como o índice de dispersão resultou em um número maior do que um indicando sobredispersão dos dados. Vale ressaltar também que a proporção de zeros é de 86%.

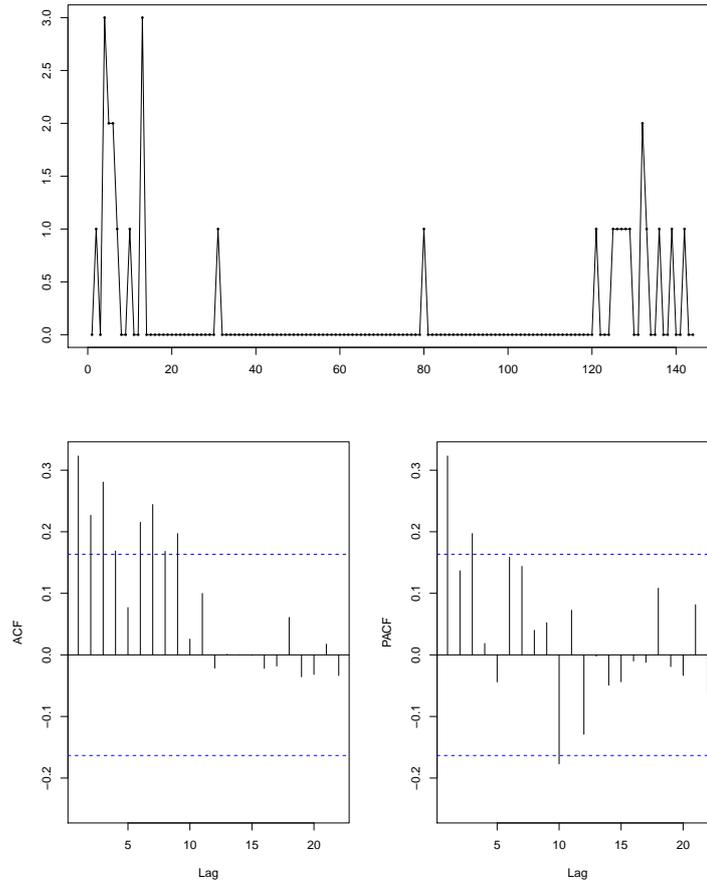


Figura 4.1: Gráficos da série, função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) para o número de pessoas que bebem em local público.

Fonte: Elaborada pelo autor

Tabela 4.1: Estatística descritiva do para o número de pessoas que bebem em local público.

Mínimo	Máximo	Média	Mediana	Variância	Índice de Dispersão	Proporção de zeros
0	3	0.1875	0	0.2793	1.4895	86%

Fonte: Elaborada pelo autor

Na Tabela 4.2 fornecemos as estimativas dos parâmetros juntamente com os seus respectivos erros padrão (entre parênteses). Para o NGINAR(1) neste banco de dados, obtêm-se as estimativas de $\hat{\alpha} = 0.1904$ e $\hat{\mu} = 0.1867$, note que a estimativa de $\hat{\alpha}$ está fora do espaço paramétrico que está definido por $\left(0, \frac{\hat{\mu}}{1+\hat{\mu}}\right) = (0, 0.1573)$. Dessa forma, o processo NGINAR(1) não é definido para esses valores dos parâmetros e, portanto, este modelo não pode ser usado para ajustar esses dados. Já o ZMGINAR(1) pode ser considerado uma vez que a estimativa do $\hat{\alpha}$ está no espaço paramétrico estabelecido, ou seja, $\left(\max\left\{0, \frac{\hat{\pi}\hat{\mu}}{1+\hat{\pi}\hat{\mu}}\right\}, \frac{\hat{\mu}}{1+\hat{\mu}}\right) = (0.1952, 0.3001)$. Para verificar a inflação de zeros de acordo

com o teste fornecido pelo Barreto-Souza (2015), testa-se a hipótese nula $H_0 : \pi = 0$ contra a hipótese alternativa $H_1 : \pi > 0$. Com base na estimativas obtidas na Tabela 4.2, temos que, $0.5655 > 0.474937 = q_{1-\gamma} \sqrt{\frac{(1+\hat{\alpha})(1+\hat{\mu})}{(n\hat{\mu}(1-\hat{\alpha}))}}$, e portanto, rejeita-se a hipótese $H_0 : \pi = 0$ em favor da hipótese alternativa $H_1 : \pi > 0$ com nível de significância de 5%, além disso é apresentado o Erro padrão (entre parênteses) calculado a partir do inverso da matriz hessiana. Dessa forma, temos evidências de que há inflação de zeros para o número de pessoas que bebem em local público. A proporção de zeros estimada pelo ZMGINAR(1) resultou em 86.9%, ou seja, próximo do valor amostral.

Tabela 4.2: Estimativa dos Parâmetros juntamente com seus respectivos erros padrão.

Modelo	Parâmetros estimados
ZMGINAR(1)	$\hat{\pi} = 0.5655$ (0.1526)
	$\hat{\alpha} = 0.2984$ (0.1333)
	$\hat{\mu} = 0.4288$ (0.1847)

Fonte: Elaborada pelo autor

Para demonstrar a aplicação dos gráficos de controle de Shewhart e CUSUM propostos para o modelo ZMGINAR(1), consideramos as 100 primeiras observações do banco de dados do número de pessoas que bebem em local público a serem usadas na fase 1 do processo, utilizadas na estimação dos parâmetros e determinação dos limites de controle e os 44 valores restantes na fase 2 do processo, para fins de monitoramento.

Tabela 4.3: Estatísticas descritivas para o número de pessoas que bebem em local público para as 100 primeiras observações.

Mínimo	Máximo	Média	Mediana	Variância	Índice de Dispersão	Proporção de zeros
0	3	0.15	0	0.2904	1.9360	91%

Fonte: Elaborada pelo autor

Na Tabela 4.3 apresenta as estatísticas descritivas a média e variância, representadas por, \bar{x} e s^2 , respectivamente. Assim a média e a variância das 100 primeiras observações são $\bar{x} = 0.15$ e $s^2 = 0.2904$, respectivamente, indicando sobredispersão dos dados, com índice de dispersão dado por $s^2/\bar{x} = 1.94$ e proporção de zeros de 91%. Na Tabela 4.4 são expostas as estimativas dos parâmetros e os respectivos erros padrão (entre parênteses) das 100 primeiras observações, note que, a estimativa do $\hat{\alpha}$ está dentro espaço paramétrico.

Com base na Tabela 4.4 testa-se a hipótese nula $H_0 : \pi = 0$ contra a hipótese alternativa $H_1 : \pi > 0$. Com base na estimativas obtidas na Tabela 4.4 pelo método de

Tabela 4.4: Estimativa dos Parâmetros juntamente com seus respectivos erros padrão das 100 primeiras observações.

Modelo	Parâmetros estimados
ZMGINAR(1)	$\hat{\pi} = 0.7159$ (0.1077)
	$\hat{\alpha} = 0.3386$ (0.1677)
	$\hat{\mu} = 0.5333$ (0.3253)

Fonte: Elaborada pelo autor

máxima de verossimilhança, temos que, $0.7159 > 0.3967 = q_{1-\gamma} \sqrt{\frac{(1+\hat{\alpha})(1+\hat{\mu})}{(n\hat{\mu}(1-\hat{\alpha}))}}$, e portanto, rejeita-se a hipótese $H_0 : \pi = 0$ em favor da hipótese alternativa $H_1 : \pi > 0$ com nível de significância de 5%. Dessa forma, temos evidências de que há inflação de zeros para o número de pessoas que bebem em local público de acordo com as 100 primeiras observações. A proporção de zeros estimada pelo ZMGINAR(1) resultou em 91%.

Em seguida, usaremos as estimativas dos parâmetros do processo obtidos na Tabela 4.4 e apresentaremos a construção e aplicação dos gráficos de controle de Shewhart e CUSUM. Considerando o processo ZMGINAR(1) com $\hat{\alpha} = 0.3386$, $\hat{\mu}_0 = 0.5333$ e $\hat{\pi} = 0.7159$, para o gráfico de Shewhart, o limite superior de controle resultou em $LSC = 4$ com um $NMA_0 = 275.150$ e para o gráfico de CUSUM com $(h, k) = (3, 1)$, resultou em $NMA_0 = 248.996$. Na Figura 4.2, apresenta-se o gráfico de Shewhart para todos os dados da amostra (Fase 1: 1-100; Fase 2: 101-142) com um limite superior de controle de $LSC = 4$, o gráfico de CUSUM com $h = 3$ e o gráfico de Saltos, este gráfico definido anteriormente representa um gráfico de Shewhart com limites escolhidos de $\pm 3\sigma_J$ como foi proposto por Weiß (2009) que definiu os limites a partir da variância dos saltos dada por $\sigma_J = \sqrt{Var(J_t)}$ definido na prova do Teorema 2.2.1 no Apêndice. De acordo com a Figura 4.2 (c) podemos observar que os resíduos não estão correlacionados e que não há um ponto de impacto que resulte na não adequação do modelo aos dados. Portanto, o modelo ZMGINAR(1) parece estar bem ajustado ao número de pessoas que bebem em local público. Já na Figura 4.2 (a), analisando o gráfico de Shewhart na Fase 1 do processo, percebe-se que este não aciona alarmes durante a Fase 1, já para o CUSUM, no caso (b), desencadeou dois alarmes. Ou seja, para a Fase 1 é adequado o gráfico de Shewhart. Já na Fase 2 do processo, os gráficos de Shewhart e CUSUM não acionaram alarmes. Ou seja, nesta análise mostrou-se que o gráfico de Shewhart é mais útil do que o gráfico de CUSUM pelo $NMA_0 = 248.996$ do gráfico CUSUM contra o $NMA_0 = 275.150$

do gráfico de Shewhart demorando-se mais a ter um alarme falso, ou seja, quanto maior o NMA_0 melhor. Então, como foi feito na simulação no cenário com inflação de zeros, o gráfico de Shewhart mostrou-se mais eficiente do que o de CUSUM.

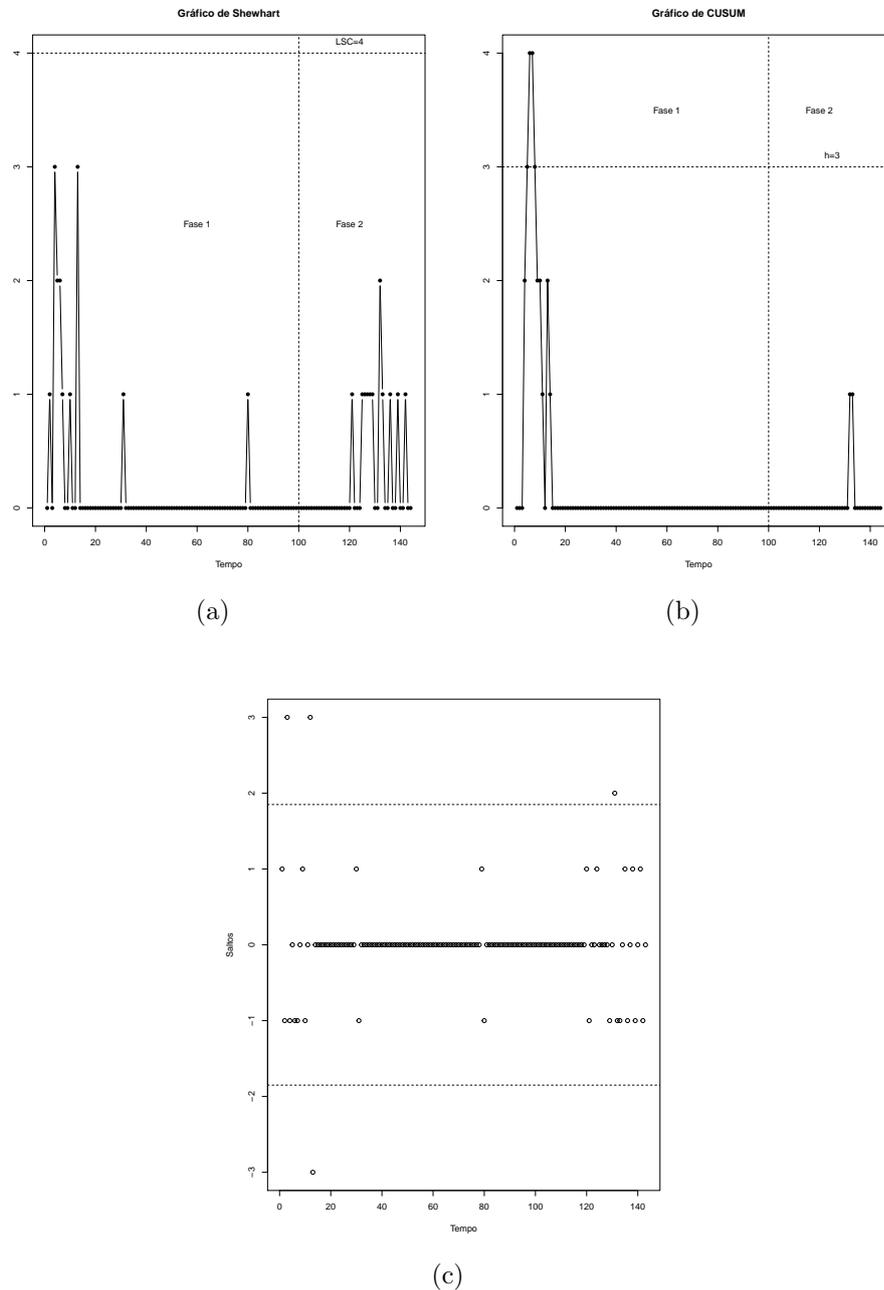


Figura 4.2: Gráficos de controle para o monitoramento do processo ZMGINAR(1) referentes ao número de pessoas que bebem em local público: (a) Gráfico de Shewhart (Fase 1: 1-100; Fase 2: 101-144), (b) Gráfico de CUSUM (Fase 1: 1-100; Fase 2: 101-144) e (c) Gráfico dos Saltos (Todas as observações).

Fonte: Elaborada pelo autor

4.2 Deflação de Zeros

O banco de dados analisado representa o número de casos de Poliomelite nos EUA, no período de Janeiro de 1970 à Novembro de 1983, disponível em <https://datamarket.com/data/set/22u4/monthly-us-polio-cases> totalizando 167 observações. A Figura 4.3 mostra o gráfico da série, a ACF e a PACF, nota-se que a utilização de modelos autorregressivos de primeira ordem é apropriada, pois no gráfico da PACF apenas a primeira defasagem se destaca dos demais, e no gráfico da ACF nota-se um decaimento exponencial. A Tabela 4.5 mostra as estatísticas descritivas dos dados.

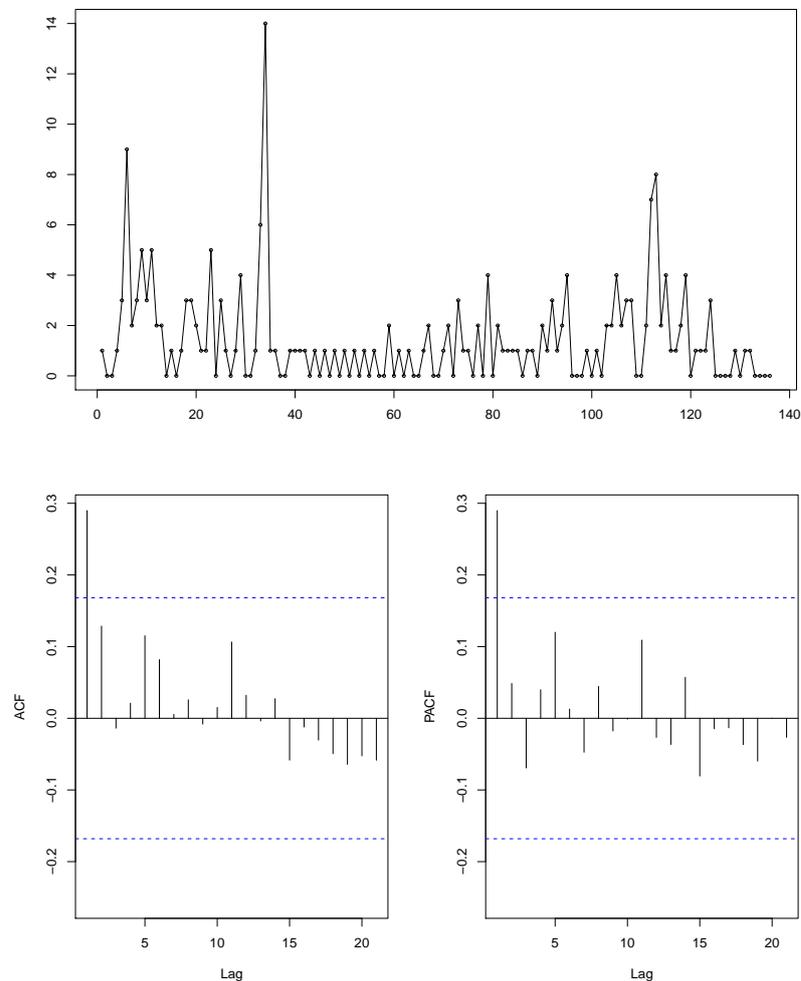


Figura 4.3: Gráficos da série, função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) para o número de casos de Poliomelite.

Fonte: Elaborada pelo autor

Pela Tabela 4.5, observa-se que a variância é maior do que a média, indicando

Tabela 4.5: Estatísticas descritivas para o número de casos de Poliomelite.

Mínimo	Máximo	Média	Mediana	Variância	Índice de Dispersão	Proporção de zeros
0	14	1.3413	1	3.5153	2.6208	37%

Fonte: Elaborada pelo autor

sobredispersão para o conjunto de dados analisados com 37% de zeros. Na Tabela 4.6 foi comparado o modelo NGINAR(1) com o modelo ZMGINAR(1). Para investigar a performance dos modelos, foram utilizados os critérios de informação de Akaike (AIC) e o critério de informação Bayesiano (BIC), que são respectivamente dados por

$$\text{AIC} = -2 \ell_{\max} + 2p, \quad \text{BIC} = -2 \ell_{\max} + p \ln(n),$$

em que ℓ_{\max} é o máximo valor da função de log-verossimilhança, p é o número de parâmetros do modelo e n é o tamanho da amostra.

Tabela 4.6: Estimativa dos parâmetros com seus respectivos erros-padrão (entre parênteses), AIC e BIC.

Modelo	Parâmetros estimados	AIC	BIC
ZMGINAR(1)	$\hat{\alpha} = 0.1894$ (0.1046)	531.8194	541.1734
	$\hat{\mu} = 1.0890$ (0.1588)		
	$\hat{\pi} = -0.2517$ (0.1044)		
NGINAR(1)	$\hat{\alpha} = 0.0830$ (0.0858)	536.3822	542.6182
	$\hat{\mu} = 1.3247$ (0.1463)		

Fonte: Elaborada pelo autor

Para verificar a deflação de zeros de acordo com o teste fornecido pelo Barreto-Souza (2015) testa-se a hipótese nula $H_0 : \pi = 0$ contra a hipótese alternativa $H_1 : \pi < 0$. Com base na estimativas obtidas na Tabela 4.6, temos que, $-0.2517 < -0.2135 = q_\gamma \sqrt{\frac{(1+\hat{\alpha})(1+\hat{\rho})}{(n\hat{\mu}(1-\hat{\alpha}))}}$, e portanto, rejeita-se a hipótese $H_0 : \pi = 0$ em favor da hipótese alternativa $H_1 : \pi < 0$ com nível de significância de 5%. Dessa forma, temos evidências de que há deflação de zeros para o número de casos de Poliomelite com respeito ao NGINAR(1). A proporção de zeros estimada pelo ZMGINAR(1) e NGINAR(1) resultaram em 34.7% e 42.9%, respectivamente, em que a proporção de zeros do ZMGINAR(1) aproximou bem da proporção de zeros dos dados. Observa-se ainda que o modelo que melhor

se ajusta aos dados foi o ZMGINAR(1), porque foi o que atingiu o valor mais baixo com relação ao AIC e BIC. Para demonstrar a aplicação dos gráficos de controle de Shewhart e CUSUM propostos para o modelo ZMGINAR(1), consideramos as 136 primeiras observações a serem usadas na fase 1 do processo, utilizadas na estimação dos parâmetros e determinação dos limites de controle e os 31 valores restantes na fase 2 do processo, para fins de monitoramento.

Tabela 4.7: Estatística Descritiva para o número de casos de Poliomelite das 136 primeiras observações.

Mínimo	Máximo	Média	Mediana	Variância	Índice de Dispersão	Proporção de zeros
0	14	1.4411	1	3.9076	2.7114	35%

Fonte: Elaborada pelo autor

A média e a variância das 136 primeiras observações são $\bar{x} = 1.4411$ e $s^2 = 3.9076$, respectivamente, indicando sobredispersão dos dados, com índice de dispersão dado por $s^2/\bar{x} = 2.7114$ e com a proporção de zeros de 35%.

Naturalmente, necessita-se comparar os modelos novamente, já que foram selecionadas as 136 primeiras observações do conjunto de dados. Na Tabela 4.8 estão os resultados.

Tabela 4.8: Estimativa dos parâmetros com seus respectivos erros-padrão (entre parênteses), AIC e BIC dos 136 primeiros casos de Poliomelite.

Modelo	Parâmetros estimados	AIC	BIC
ZMGINAR(1)	$\hat{\alpha} = 0.1722$ (0.0993) $\hat{\mu} = 1.1724$ (0.1818) $\hat{\pi} = -0.2432$ (0.1077)	447.9222	456.6602
NGINAR(1)	$\hat{\alpha} = 0.0988$ (0.0944) $\hat{\mu} = 1.4155$ (0.1731)	451.7951	457.6204

Fonte: Elaborada pelo autor

Novamente, o modelo que melhor se ajusta aos dados foi o ZMGINAR(1), porque foi o que atingiu os valores mais baixos com relação a AIC e BIC, portanto se ajustando melhor a esses dados com deflação de zeros das 136 primeiras observações do número de casos de Poliomelite. Com base na Tabela 4.8 testa-se a hipótese nula $H_0 : \pi = 0$ contra a hipótese alternativa $H_1 : \pi < 0$. Com base na estimativas obtidas na Tabela 4.8 pelo

método de máxima de verossimilhança, temos que, $-0.2432 < -0.228 = q_\gamma \sqrt{\frac{(1+\hat{\alpha})(1+\hat{\rho})}{(n\hat{\mu}(1-\hat{\alpha}))}}$, e portanto, rejeita-se a hipótese $H_0 : \pi = 0$ em favor da hipótese alternativa $H_1 : \pi < 0$ com nível de significância de 5%. Dessa forma, temos evidências de que há deflação de zeros para o número de casos de Poliomelite acordo com as 136 primeiras observações em relação ao NGINAR(1). A proporção de zeros do ZMGINAR(1) resultou em 32.9% e NGINAR(1) de 41.3%.

Em seguida, usaremos as estimativas dos parâmetros do processo e demonstraremos a construção e aplicação dos gráficos de controle de Shewhart e CUSUM. Então, para o processo ZMGINAR(1) com $\hat{\alpha} = 0.1722$, $\hat{\mu}_0 = 1.1724$ e $\hat{\pi} = -0.2432$, para o gráfico de Shewhart usando as estimativas da Tabela 4.8, o limite superior de controle resultou em $LSC = 10$ com um $NMA_0 = 387.837$, assim como o gráfico de CUSUM com $(h, k) = (10, 3)$ e $NMA_0 = 420.305$. Deve-se notar que os desempenhos dos parâmetros nos gráficos no modelo foram escolhidos a fim de obter-se um desempenho comparável entre ambos. Assim, os procedimentos do gráfico de Shewhart e CUSUM testou-se parametrizações dos gráficos de um determinado intervalo de parâmetros e selecionamos o que melhor corresponde ao NMA_0 . Portanto, para o gráfico de CUSUM, usamos a regra de Weiß e Testik (2009) para escolher o k (ver seção 4.3), e então procurou-se o inteiro h apropriado. Na Figura 4.4, apresenta-se o gráfico de Shewhart para todos os dados da amostra (Fase 1: 1-136; Fase 2: 137-167) com um limite superior de controle de $LSC = 10$, o gráfico de CUSUM com $h = 10$ e o gráfico de Saltos, este gráfico definido anteriormente representa um gráfico de Shewhart com limites escolhidos de $\pm 3\sigma_J$ como foi proposto por Weiß (2009), defini-se os limites da seguinte forma, a variância dos saltos dada por $\sigma_J = \sqrt{Var(J_t)}$ definido na prova do Teorema 2.2.1 no Apêndice. De acordo com a Figura 4.4 (c), os resíduos não estão correlacionados e que não há um ponto de impacto no modelo. Portanto, o modelo ZMGINAR(1) parece estar bem ajustado ao número casos de Poliomelite. Já na Figura 4.4 (a), analisando o uso do gráfico de Shewhart na Fase 1 do processo, percebe-se que o gráfico aciona um alarme durante a Fase 1, já para o CUSUM, no caso (b), desencadeou 2 alarmes. Ou seja, para a Fase 1 do processo recomenda-se o uso do gráfico de Shewhart. Já na Fase 2 do processo, o gráfico de Shewhart e CUSUM não acionam alarmes. Ou seja, nesta análise mostrou-se que o gráfico de CUSUM é mais útil do que o gráfico de Shewhart, além de que pelo $NMA_0 = 420.305$ do gráfico CUSUM contra o $NMA_0 = 387.837$ do gráfico de Shewhart, mostra que com o CUSUM, demora-se

mais a ter um alarme falso, ou seja, quanto maior o NMA_0 melhor. Então, como foi feito na simulação no cenário com deflação de zeros, o gráfico de CUSUM mostrou-se mais eficiente em termos de NMA_0 em relação ao gráfico de Shewhart.

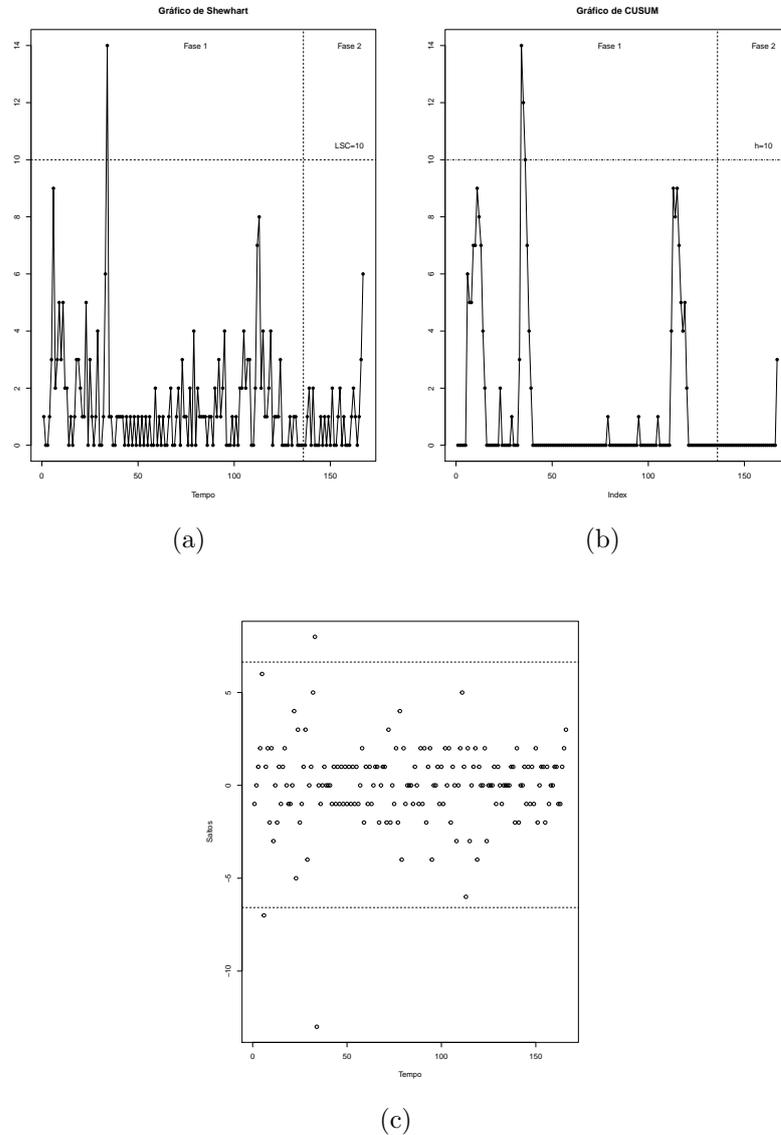


Figura 4.4: Gráficos de controle para o monitoramento do processo ZMGINAR(1) referentes ao número casos de Poliomelite nos EUA: (a) Gráfico de Shewhart (Fase 1: 1-136; Fase 2: 136-167), (b) Gráfico de CUSUM (Fase 1: 1-136; Fase 2: 136-167) e (c) Gráfico dos Saltos (Todas as observações).

Fonte: Elaborada pelo autor

Capítulo 5

Considerações Finais

As principais contribuições deste trabalho referem-se à proposição de novos estimadores do modelo ZMGINAR(1) e ao desempenho dos gráficos de Shewhart e CUSUM na detecção de deslocamento positivo da média do processo ZMGINAR(1) introduzido por Barreto-Souza (2015) sendo avaliado pelo critério do NMA nos gráficos apresentados.

A princípio foram apresentados novos estimadores para o modelo ZMGINAR(1) como o estimador de MV, YW e o estimador PB e comparamos os estimadores em termos de viés e EQM com o estimador MQC proposto por Barreto-Souza (2015). A partir do estudo simulado dos estimadores propostos apresentado no Capítulo 2, notou-se, em geral, em termos de EQM e viés que os estimadores propostos apresentaram um desempenho melhor para todos os parâmetros do modelo.

No Capítulo 3, apresentamos os gráficos de controle de Shewhart e CUSUM a serem aplicado ao modelo ZMGINAR(1). Utilizando uma técnica apropriada de cadeia de Markov, avaliamos o desempenho dos NMA a partir das configurações apresentadas. O NMA_0 foi considerado como a medida de desempenho apropriada quando o processo está em controle, enquanto que o desempenho fora de controle foi avaliado em termos de NMA_1 . Foram também fornecidas diretrizes práticas para a seleção dos valores para os parâmetros de concepção de cada configuração. Os resultados mostraram que a inflação ou a deflação de zeros requer um ajuste para os valores dos parâmetros cada gráfico para preservar as taxas de falso alarme para os níveis desejados. Nas simulações de inflação de zeros, notou-se quando a proporção de zeros é alta, aproximadamente, acima de 40%, o gráfico de controle de Shewhart é melhor comparado ao CUSUM, já para proporções menores o CUSUM é mais eficiente. Dessa forma, nas simulações de deflação de zeros a

medida que o π diminuía negativamente, o CUSUM se mostrava mais eficiente do que o de Shewhart em alguns cenários, mesmo com a proporção de zeros sendo alta.

Por fim, nas aplicações com inflação e deflação de zeros, os resultados mostraram que na inflação de zeros o gráfico de Shewhart é melhor em relação ao CUSUM, em termos de NMA_0 , já na deflação de zeros o CUSUM tornou-se mais eficiente do que o Shewhart em termos de NMA_0 .

Apêndice

Propriedades do operador *thinning binomial negativo*

- Prova do Lema 2.2.1

i) $E[\alpha * X] = \alpha E[X]$.

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

□

ii) $E[\alpha * X|X] = \alpha X$.

$$E[\alpha * X|X] = E\left[\sum_{i=1}^X G_i|X\right] = \sum_{i=1}^X E(G_i|X) = \sum_{i=1}^X E(G_i) = \alpha X.$$

□

iii) $Var[\alpha * X|X] = \alpha(1 + \alpha)X$

$$\begin{aligned} Var[\alpha * X|X] &= Var\left(\sum_{i=1}^X G_i|X\right) = \sum_{i=1}^X Var(G_i|X) \quad (\text{Independência}) \\ &= \sum_{i=1}^X Var(G_i) \quad G_i \sim \text{Geométrica}\left(\frac{\alpha}{1 + \alpha}\right) = X[\alpha(1 + \alpha)] = \alpha(1 + \alpha)X. \end{aligned}$$

□

$$\text{iv) } \text{Var}[\alpha * X] = \alpha^2 \text{Var}[X] + \alpha(1 + \alpha)E[X]$$

$$\begin{aligned} \text{Var}[\alpha * X] &= \underbrace{\text{Var}(E(\alpha * X|X))}_{\text{propriedade. ii}} + E[\underbrace{\text{Var}(\alpha * X|X)}_{\text{propriedade. iii}}] \\ &= \text{Var}[\alpha X] + E[\alpha(1 + \alpha)X] \\ &= \alpha^2 \text{Var}[X] + \alpha(1 + \alpha)E[X] \end{aligned}$$

□

$$\text{v) } E[\alpha * X]^2 = \alpha(1 + \alpha)E[X] + \alpha^2 E[X^2].$$

$$\begin{aligned} E[\alpha * X]^2 &= E[E[(\alpha * X)^2|X]] \\ &= E\left[E\left[\left(\sum_{i=1}^X \alpha * X\right)^2 \middle| X\right]\right] = E\left[E\left[\left(\sum_{i=1}^X G_i^2 + \sum_{j=1}^X \sum_{\substack{k=1 \\ j \neq k}}^X G_j G_k\right) \middle| X\right]\right] \\ &= E\left[\sum_{i=1}^X [G_i^2|X] + \sum_{j=1}^X \sum_{\substack{k=1 \\ j \neq k}}^X E[G_j G_k|X]\right] = E\left[\sum_{i=1}^X (G_i^2) + \sum_{j=1}^X \sum_{\substack{k=1 \\ j \neq k}}^X E(G_j G_k)\right] \\ &= E[X(\alpha(1 + \alpha) + \alpha^2) + X(X - 1)\alpha\alpha] = E[\alpha(1 + \alpha)X + \alpha_1^2 X + \alpha^2(X^2 - X)] \\ &= E[\alpha(1 + \alpha)X + \alpha^2 X + \alpha^2 X^2 - \alpha^2 X] = E[\alpha(1 + \alpha)X] + E[\alpha^2 X^2] = \alpha(1 + \alpha)E[X] + \alpha^2 E[X^2]. \end{aligned}$$

□

$$\text{vi) } E[\alpha * X]^3 = \alpha E[X^3] + 3\alpha^2[1 + \alpha]E[X^2] + \alpha(1 + \alpha)(1 + 2\alpha)E[X].$$

$$\begin{aligned} E[\alpha * X]^3 &= E \left[E [(\alpha * X)^3 | X] \right] \\ &= E \left[E \left[\sum_{i=1}^X G_{i,1}^3 + 3 \sum_{i=1}^X \sum_{\substack{j=1 \\ i \neq j}}^X G_{i,1}^2 G_{j,1} + \sum_{i=1}^X \sum_{\substack{j=1 \\ i \neq j}}^X \sum_{\substack{k=1 \\ i \neq k \\ j \neq k \\ i \neq j}}^X G_{i,1} G_{j,1} G_{k,1} \middle| X \right] \right] \\ &= E \left[\sum_{i=1}^X E[G_{i,1}^3 | X] + 3 \sum_{i=1}^X \sum_{\substack{j=1 \\ i \neq j}}^X E[G_{i,1}^2 G_{j,1} | X] + \sum_{i=1}^X \sum_{\substack{j=1 \\ i \neq j}}^X \sum_{\substack{k=1 \\ i \neq k \\ j \neq k \\ i \neq j}}^X E[G_{i,1} G_{j,1} G_{k,1} | X] \right] \\ &= E \left[\sum_{i=1}^X E[G_i^3] + 3 \sum_{i=1}^X \sum_{\substack{j=1 \\ i \neq j}}^X E[G_i^2 G_j] + \sum_{i=1}^X \sum_{\substack{j=1 \\ i \neq j}}^X \sum_{\substack{k=1 \\ i \neq k \\ j \neq k \\ i \neq j}}^X E[G_i G_j G_k] \right] \\ &= E[X(6\alpha^3 + 6\alpha^2 + \alpha) + 3X(X-1)[\text{Var}(G_i) + [E(G_i)]^2]E(G_j) + \\ &\quad + X(X-1)(X-2)] \\ &= E[X(6\alpha^3 + 6\alpha^2 + \alpha) + 3X(X-1)[\alpha(1+\alpha) + \alpha^2] + X(X-1)(X-2)] \\ &= E[6\alpha^3 X + 6\alpha^2 X + \alpha X + 3\alpha^2 X^2 + 6\alpha^3 X^2 - 3\alpha^2 X - 6X\alpha^3 + \\ &\quad + \alpha^3 X^3 - 3\alpha^3 X^2 + 2\alpha^3 X] \\ &= E[\alpha^3 X^3 + 3\alpha^2 X^2 + 3\alpha^3 X^2 + \alpha X + 3\alpha^2 X + 2\alpha^3 X] \\ &= \alpha E[X^3] + 3\alpha^2[1 + \alpha]E[X^2] + \alpha(1 + \alpha)(1 + 2\alpha)E[X]. \end{aligned}$$

□

$$\text{vii) } E[\prod_{i=1}^r (\alpha_i * X_i)] = \prod_{i=1}^r \alpha_i E \left[\prod_{i=1}^r X_i \right], r \geq 1.$$

$$\begin{aligned} E \left[\prod_{i=1}^r (\alpha_i * X_i) \right] &= E \left[E \left[\prod_{i=1}^r (\alpha_i * X_i) \middle| \sigma(X_1, \dots, X_r) \right] \right] \\ &= E \left[E \left[\prod_{i=1}^r \left(\sum_{i=1}^X G_i \right) \middle| \sigma(X_1, \dots, X_r) \right] \right] \\ &= \left[\prod_{i=1}^r \sum_{i=1}^r E(G_i) \right] = E \left[\prod_{i=1}^r X \alpha_i \right] = \prod_{i=1}^r \alpha_i E \left[\prod_{i=1}^r X_i \right], r \geq 1. \end{aligned}$$

Prova do Teorema 2.2.1

Demonstração. Suponha $\{X_t\}_{t=0}^{\infty}$ um processo estacionário ZMGINAR(1). Seja φ_{X_t} e φ_{ϵ_t} as funções geradoras de probabilidade das variáveis X_t e ϵ_t . Considere o processo bivariado (X_t, X_{t-1}) . Como $\varphi_{X_t}(s) = \frac{1+\pi\mu(1-s)}{1+\mu(1-s)}$ e $\varphi_{\epsilon_t}(s) = \frac{1+\alpha(1+\mu)(1-s)}{1+\mu(1-s)} \frac{1+\pi\mu(1-s)}{1+\alpha(1+\pi\mu)(1-s)}$, logo

$$\begin{aligned} \varphi_{X_t, X_{t-1}}(u, v) &= E[u^{X_t} v^{X_{t-1}}] = E[u^{\alpha^* X_{t-1} + \epsilon_t} v^{X_{t-1}}] = E[u^{\epsilon_t} v^{X_{t-1}} E(u^{\alpha^* X_{t-1}} | X_{t-1})] \\ &= \varphi_{\epsilon_t}(u) E \left[v^{X_{t-1}} \left(\frac{1}{1 + \alpha - \alpha u} \right)^{X_{t-1}} \right] = \varphi_{\epsilon_t}(u) \varphi_{X_t} \left(\frac{\mu}{1 + \alpha - \alpha u} \right) \\ &= \varphi_{\epsilon_t}(u) \frac{1 + \pi\mu \left(1 - \frac{\mu}{1 + \alpha - \alpha u}\right)}{1 + \mu \left(1 - \frac{\mu}{1 + \alpha - \alpha u}\right)} = \varphi_{\epsilon_t}(u) \frac{1 + \pi\mu[1 + \alpha(1 - u)] - \pi\mu v}{1 + \mu[1 + \alpha(1 - u)] - \mu v}. \end{aligned}$$

□

Seja φ_{J_t} a função geradora de probabilidade de momentos da variável J_t , em que $J_t = X_t - X_{t-1}$, $t \geq 2$, logo, partindo do Teorema 2.2.1

$$\begin{aligned} \varphi_{J_t}(\mu) &= E[e^{v J_t}] = E[e^{v X_t} e^{-v X_{t-1}}] = \varphi_{X_{t-1}, X_t}(v^{-1}, v) \\ &= \varphi_{\epsilon_t}(v) \frac{1 + \pi\mu[1 + \alpha(1 - v)] - \pi\mu/v}{1 + \mu[1 + \alpha(1 - v)] - \mu/v}. \end{aligned}$$

Como $E(J_t^n) = \varphi_{J_t}^n(1)$, obtemos $E(J_t) = 0$ e $\sigma_J = \text{Var}(J_t) = E[J_t - E(J_t)]^2 = E(J_t^2) = 2\mu(1 - \alpha)(1 - \pi)[1 + \mu(1 + \pi)]$. A função de autocorrelação dos saltos (J_t) é dado por:

$$\begin{aligned} \rho_J(k) &= \frac{\text{Cov}(J_{t+k}, J_t)}{\text{Var}(J_t)} = \frac{1}{\text{Var}(J_t)} \text{Cov}(X_{t+k} - X_{t+k-1}, X_t - X_{t-1}) \\ &= \frac{1}{\text{Var}(J_t)} \text{Var}(X_t) [2\rho(k) - \rho(k+1) - \rho(k-1)] \\ &= \frac{1}{2\mu(1 - \alpha)(1 - \pi)[1 + \mu(1 + \pi)]} \mu(1 - \pi)[1 + \mu(1 + \pi)] (2\rho(k) - \rho(k+1) - \rho(k-1)) \\ &= \frac{(2\alpha^k - \alpha^{k+1} - \alpha^{k-1})}{2(1 - \alpha)} \\ &= \frac{-\alpha^{k+1}(\alpha^2 - 2\alpha + 1)}{2(1 - \alpha)} = \frac{-\alpha^{k+1}(\alpha - 1)^2}{2(1 - \alpha)} = -\alpha^{k-1} \left(\frac{1 - \alpha}{2} \right), \quad \text{para } k \in \mathbb{N}^*. \end{aligned}$$

em que $\rho(\cdot)$ é a função de autocorrelação de X_t . Sendo que a autocorrelação entre X_{t+k}

e X_t é dada por

$$\rho(k) = \text{corr}(X_{t+k}, X_t) = \alpha^k.$$

Note que, a função de autocorrelação dos saltos é sempre negativa.

Referências Bibliográficas

AHN, S.; LEE, G.; JEON, J. Analysis of the M/D/1-type queue based on an integer-valued first-order autoregressive process. *Operations Research Letters*, Elsevier, v. 27, n. 5, p. 235–241, 2000. 6

AL-OSH, M. A.; ALZAID, A. A. First-Order Integer-Valued Autoregressive (INAR(1)) Process. *Journal of Time Series Analysis*, Wiley Online Library, v. 8, n. 3, p. 261–275, 1987. 2, 6

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*, Wiley Online Library, v. 36, n. 6, p. 839–852, 2015. 2, 3, 4, 5, 7, 8, 9, 10, 13, 14, 15, 18, 38, 42, 46

BOURGUIGNON, M.; WEIß, C. H. An INAR(1) process for modeling count time series with equidispersion, underdispersion and overdispersion. *TEST*, Springer, v. 1, n. 1, p. 1–22, 2017. 2, 17

BROOK, D.; EVANS, D. An Approach to the Probability Distribution of CUSUM Run Length. *Biometrika*, JSTOR, v. 59, n. 3, p. 539–549, 1972. 25, 26

CHAO-WEN, L.; JR, M. R. R. CUSUM Charts For Monitoring An Autocorrelated Process. *Journal of Quality Technology*, American Society for Quality, v. 33, n. 3, p. 316, 2001. 25

FERLAND, R.; LATOUR, A.; ORAICHI, D. Integer-Valued GARCH Process. *Journal of Time Series Analysis*, Wiley Online Library, v. 27, n. 6, p. 923–942, 2006. 6

GOMES, K. S. Modelagem INAR(p) para previsão de índices de qualidade do ar. UFES, v. 1, n. 1, p. 1–71, 2009. Dissertação de Mestrado. 6

HAWKINS, D. M.; OLWELL, D. H. *Cumulative Sum Charts and Charting for Quality Improvement*. [S.l.]: Springer Science & Business Media, 2012. v. 1. 1-246 p. 25, 27

HE, S.; HUANG, W.; WOODALL, W. H. CUSUM charts for monitoring a zero-inflated poisson process. *Quality and Reliability Engineering International*, Wiley Online Library, v. 28, n. 2, p. 181–192, 2012. 2

JAZI, M. A.; JONES, G.; LAI, C.-D. First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis*, Wiley Online Library, v. 33, n. 6, p. 954–963, 2012. 2

- JOHANSSON, P. Speed limitation and motorway casualties: a time series count data regression approach. *Accident Analysis & Prevention*, Elsevier, v. 28, n. 1, p. 73–87, 1996. 6
- LI, C.; WANG, D.; ZHU, F. Effective Control Charts for Monitoring the NGINAR(1) Process. *Quality and Reliability Engineering International*, Wiley-Blackwell, v. 32, n. 3, p. 877–888, 2016. 2, 22, 23
- LOPES, T. L. d. C. *Novos Modelos para séries temporais de valores binários e inteiros não negativos baseados em operadores thinning*. UFRN, 1–84 p. Dissertação (Mestrado), 2016. 6
- LUCAS, J. M.; CROSIER, R. B. Fast Initial Response for CUSUM Quality-Control Schemes: Give Your CUSUM A Head Start. *Technometrics*, Taylor & Francis, v. 24, n. 3, p. 199–205, 1982. 25
- MCKENZIE, E. Some simple models for discrete variate time series. *JAWRA Journal of the American Water Resources Association*, Wiley Online Library, v. 21, n. 4, p. 645–650, 1985. 2
- MEGAHED, F. M. et al. A Note on the ARL of two-sided Bernoulli-based CUSUM Control Charts. *Journal of Quality Technology*, American Society for Quality, v. 43, n. 1, p. 43, 2011. 25
- MONTGOMERY, D. C. *Introduction to Statistical Quality Control*. [S.l.]: John Wiley & Sons, 2007. v. 6. 1–734 p. 1, 22
- OVIDO-TRESPALACIOS, O. et al. The Application of the Markov Chain in Statistical Quality Monitoring. v. 1, p. 68–72, 2014. 25
- PAGE, E. Continuous Inspection Schemes. *Biometrika*, JSTOR, v. 41, n. 1/2, p. 100–115, 1954. 24
- PEREIRA, M. B. *Modelos inar sazonais e de raízes unitárias*. UFPE, 1–68 p. Dissertação (Mestrado), 2011. 6
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>. 16, 28
- RAKITZIS, A. C.; CASTAGLIOLA, P. The effect of parameter estimation on the performance of one-sided shewhart control charts for zero-inflated processes. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 45, n. 14, p. 4194–4214, 2016. 2
- RAKITZIS, A. C.; CASTAGLIOLA, P.; MARAVELAKIS, P. E. Cumulative sum control charts for monitoring geometrically inflated poisson processes: An application to infectious disease counts data. *Statistical methods in medical research*, SAGE Publications Sage UK: London, England, v. 1, n. 1, p. 1–19, 2016. 2
- RAKITZIS, A. C.; MARAVELAKIS, P. E.; CASTAGLIOLA, P. CUSUM control charts for the Monitoring of Zero-inflated Binomial Processes. *Quality and Reliability Engineering International*, Wiley Online Library, v. 32, n. 2, p. 465–483, 2016. 2

- RAKITZIS, A. C.; WEIß, C. H.; CASTAGLIOLA, P. Control Charts for Monitoring Correlated Poisson Counts with an Excessive Number of Zeros. *Quality and Reliability Engineering International*, Wiley Online Library, v. 33, n. 2, p. 413–430, 2016. 2, 22, 23, 24, 25, 28
- RAKITZIS, A. C.; WEIß, C. H.; CASTAGLIOLA, P. Control charts for monitoring correlated counts with a finite range. *Applied Stochastic Models in Business and Industry*, Wiley Online Library, 2017. 2
- RISTIC, M. M.; BAKOUCH, H. S.; NASTIC, A. S. A New Geometric First-Order Integer-Valued Autoregressive (NGINAR(1)) Process. *Journal of statistical planning and inference*, Elsevier, v. 139, n. 7, p. 2218–2226, 2009. 2, 8, 10
- SAMOHYL, R. W. *Controle Estatístico de Qualidade*. [S.l.]: Elsevier, 2009. v. 1. 1–274 p. 24
- SHEWHART, W. The application of statistics as an aid in maintaining quality of a manufactured product. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 20, n. 152, p. 546–548, 1925. 23
- SILVA, F. R. d. *Estimação e previsão no processo INARCH (2)*. 1–91 p. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2016. 6
- SILVA, M. E. D.; OLIVEIRA, V. L. Difference equations for the higher-order moments and cumulants of the INAR(1) model. *Journal of Time Series Analysis*, Wiley Blackwell, v. 25, n. 3, p. 317–333, 2004. 8
- STEUTEL, F.; HARN, K. V. Discrete Analogues of Self-Decomposability and Stability. *The Annals of Probability*, JSTOR, v. 7, n. 5, p. 893–899, 1979. 6
- WEIß, C. The markov chain approach for performance evaluation of control charts a tutorial. *Process control: Problems, techniques and applications*, p. 205–228, 2011. 23
- WEIß, C. H. The combined INAR (p) models for time series of counts. *Statistics & Probability Letters*, Elsevier, v. 78, n. 13, p. 1817–1822, 2008. 6
- WEIß, C. H. Jumps in binomial ar (1) processes. *Statistics & Probability Letters*, Elsevier, v. 79, n. 19, p. 2012–2019, 2009. 39, 44
- WEIß, C. H. Detecting Mean Increases in Poisson INAR(1) Processes with EWMA Control Charts. *Journal of Applied Statistics*, Taylor & Francis, v. 38, n. 2, p. 383–398, 2011. 2
- WEIß, C. H.; TESTIK, M. C. CUSUM Monitoring of First-Order Integer-Valued Autoregressive Processes of Poisson Counts. *Journal of Quality Technology*, American Society for Quality, v. 41, n. 4, p. 389, 2009. 2, 24, 25, 28, 44
- WEIß, C. H.; TESTIK, M. C. The Poisson INAR(1) CUSUM chart under overdispersion and estimation error. *IIE Transactions*, Taylor & Francis, v. 43, n. 11, p. 805–818, 2011. 2
- WEIß, C. H.; TESTIK, M. C. Detection of Abrupt Changes in Count Data Time Series: Cumulative Sum Derivations for INARCH(1) Models. *Journal of quality technology*, American Society for Quality, v. 44, n. 3, p. 249, 2012. 24, 25

ZHENG, H.; BASAWA, I. V. First-order observation-driven integer-valued autoregressive processes. *Statistics & Probability Letters*, Elsevier, v. 78, n. 1, p. 1–9, 2008. 6

ZHOU, J.; BASAWA, I. Least-squares estimation for bifurcating autoregressive processes. *Statistics & probability letters*, Elsevier, v. 74, n. 1, p. 77–88, 2005. 6