

# Modelos para Séries Temporais Utilizando as Distribuições Normal Generalizada e Log-Normal Generalizada

Eder Angelo Milani

# Modelos para Séries Temporais Utilizando as Distribuições Normal Generalizada e Log-Normal Generalizada

Eder Angelo Milani

Orientador: Prof. Dr. Marinho G. Andrade

Tese apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística.

UFSCar - São Carlos

Maio/2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar  
Processamento Técnico  
com os dados fornecidos pelo(a) autor(a)

M637m Milani, Eder Angelo  
Modelos para séries temporais utilizando as distribuições normal generalizada e log-normal generalizada / Eder Angelo Milani. -- São Carlos : UFSCar, 2016.  
139 p.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2016.

1. Modelo Normal Generalizada-ARMA. 2. Modelo Box-Cox Normal Generalizada-ARMA. 3. Modelo Normal Generalizada-PAR. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Estatística

---

Folha de Aprovação

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de tese de doutorado do candidato Eder Angelo Milani, realizada em 23/03/2016:

---

Prof. Dr. Marinho Gomes de Andrade Filho  
USP

---

Prof. Dr. Mauricio Enrique Zevallos Herencia  
UNICAMP

---

Prof. Dr. Rafael Izbicki  
UFSCar

---

Profa. Dra. Thelma Sáfadi  
UFLA

---

Prof. Dr. Valderio Anselmo Reisen  
UFES

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Prof. Dr. Valderio Anselmo Reisen e, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do aluno Eder Angelo Milani.

---

Prof. Dr. Marinho Gomes de Andrade Filho  
Presidente da Comissão Examinadora  
USP

# Agradecimentos

Agradeço primeiramente à Deus por me dar forças para alcançar o final de mais uma etapa da minha vida.

À minha mãe, Maria, ao meu pai, José, à minha irmã, Eliana, ao meu cunhado, Carlos, e à minha linda sobrinha, Isabella, que me deram todo o apoio e incentivo para que eu chegasse até esta etapa de minha vida.

À minha namorada, Amanda Buosi Gazon, por ser companheira, amiga, paciente e por não medir esforços para me ajudar nesta conquista.

Ao Prof. Dr. Marinho G. Andrade por dividir comigo seus conhecimentos e se tornar um amigo que nunca mais irei esquecer.

Aos Professores Dr. Ricardo S. Ehlers, Dr. Valderio A. Reisen, Dr. Maurício E. Z. Herencia, Dr. Rafael Izbicki, Dra. Thelma Sáfydi pelas sábias sugestões e comentários que foram de fundamental importância para alcançar o final deste trabalho.

Aos Professores do Departamento de Estatística que dividiram comigo todos os sentimentos que o doutorado me fez sentir.

Aos meus amigos e colegas que me acompanharam durante a realização deste trabalho, em especial, Willian Oliveira, Vinicius e Marcos William.

A todos do departamento de estatística que contribuíram de maneira direta ou indireta para a minha formação.

À Capes pelo apoio financeiro.

# Resumo

A partir da distribuição normal generalizada e dos conceitos do modelo autorregressivo e de médias móveis generalizado, introduzimos o modelo normal generalizada-ARMA, como alternativa para modelar séries temporais, que exibem simetria e caudas mais leves ou mais pesadas quando comparadas com a distribuição normal. Apresentamos aplicações do modelo proposto, usando três séries temporais, das áreas de hidrologia, políticas públicas e economia. O modelo proposto se apresentou como uma boa alternativa ao modelo ARMA com distribuição normal. Estendemos o modelo para o caso de séries que apresentam assimetria. Neste caso, utilizamos a transformação de Box-Cox, denotado por Box-Cox normal generalizada-ARMA. O caso particular quando utilizamos a transformação logarítmica é chamado de log-normal generalizada-ARMA. Ajustamos os modelos com transformação à séries de vazões das usinas hidrelétricas de Furnas e Sobradinho. Calculamos predições, que para o modelo com transformação, foram melhores, quando comparado ao modelo sem transformação. Com o objetivo de tratar séries que apresentam periodicidade na função de correlação, definimos três extensões do modelo autorregressivo periódico, chamando-os de modelo normal generalizada autorregressivo periódico, modelo log-normal generalizada autorregressivo periódico e modelo Box-Cox normal generalizada autorregressivo periódico. Constatamos que as séries de vazões das usinas hidrelétricas de Furnas e Sobradinho apresentam correlação periódica. Apresentamos duas aplicações dos modelos periódicos propostos usando estas séries. Nos ajustes dos modelos, notamos que não há necessidade da utilização da distribuição normal generalizada em todos os meses, mas em alguns a distribuição normal generalizada se sobressaiu em relação a distribuição normal. Por último, definimos a distribuição normal generalizada zero inflacionada e o modelo para séries temporais normal generalizada zero inflacionada-ARMA. Adotando o método de máxima verossimilhança e o modelo para séries que apresentam inflação de

zeros, analisamos a série da quantidade de precipitação pluviométrica da cidade de São Carlos.

**Palavras-chave:** Modelo Normal Generalizada-ARMA. Modelo Box-Cox Normal Generalizada-ARMA. Modelo Normal Generalizada-PAR. Modelo Box-Cox Normal Generalizada-PAR. Modelo Normal Generalizada Zero Inflacionada-ARMA.

# Abstract

From the generalized normal distribution and concepts of the generalized autoregressive moving averages models we introduce the generalized normal-ARMA model as an alternative way to model time series exhibiting symmetry and tails that may be lighter or heavier when compared the normal distribution. We present application for proposed model using three time series in the hydrology, economy and public policy areas. The proposed model is presented as good alternative when compared to ARMA model with normal distribution. We extended this model the case of the asymmetric time series. In this case we used the Box-Cox transformation, denoted by Box-Cox generalized normal ARMA. The particular case, when we use the logarithmic transformation is called generalized log-normal ARMA. We adjusted the models with transformation to the series on monthly average affluent streamflow of the Furnas and Sobradinho hydroelectric plants. We obtain the prediction values for the model with transformation, that are better when compared with the model without transformation. To treat time series that exhibit periodic in the correlation function we defined three extensions for periodic autoregressive model, called generalized normal periodic autoregressive model, generalized log-normal periodic autoregressive model and Box-Cox generalized normal periodic autoregressive model. We can observed that the series on monthly average affluent streamflow of the Furnas and Sobradinho hydroelectric plants have periodic correlation. We present two applications of periodic models from these series. In the models, we note that is not necessary the use of generalized normal distribution in every months, just in some the generalized normal distribution presented better results than the normal distribution. Finally, we define the generalized normal zero inflated distribution and the generalized normal zero inflated ARMA model for time series. Adopting the model for series that have zero inflation and the maximum likelihood method for estimation of parameters, we



analyze the serie of the amount of rainfall in the city of São Carlos.

**Keywords:** Generalized normal ARMA model. Box Cox Generalized Normal ARMA Model. Generalized Normal PAR Model. Box-Cox Generalized Normal PAR Model. Generalized Normal Zero Inflated ARMA Model.

# Sumário

<b>Agradecimentos</b>	<b>i</b>
<b>Resumo</b>	<b>ii</b>
<b>Abstract</b>	<b>iv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Modelo ARMA . . . . .	4
1.2 Modelo GARMA . . . . .	5
1.3 Série Temporal Periódica . . . . .	6
1.4 Proposta da Tese . . . . .	9
1.5 Organização do Texto . . . . .	10
<b>2 Modelo Normal Generalizada ARMA</b>	<b>12</b>
2.1 Distribuição Normal Generalizada . . . . .	13
2.2 Modelo Normal Generalizada ARMA . . . . .	15
2.2.1 Estimação de Máxima Verossimilhança . . . . .	19
2.2.2 Previsão . . . . .	20
2.2.3 Estudo de Simulação . . . . .	21
2.3 Aplicações . . . . .	28
2.3.1 Série de Vazões da Usina Hidrelétrica de Furnas . . . . .	28

2.3.2	Série da Média Mensal de Passageiros de Ônibus da Cidade de Iowa	33
2.3.3	Série do Índice HICP dos Estados Unidos . . . . .	38
2.4	Conclusão . . . . .	43
<b>3</b>	<b>Modelo Assimétrico Normal Generalizada ARMA</b>	<b>44</b>
3.1	Modelo Log-Normal Generalizada ARMA . . . . .	45
3.1.1	Previsão . . . . .	46
3.1.2	Aplicação à Série de Furnas . . . . .	47
3.2	Modelo Box-Cox Normal Generalizada ARMA . . . . .	50
3.2.1	Estimação de Máxima Verossimilhança . . . . .	51
3.2.2	Previsão . . . . .	53
3.2.3	Estudo de Simulação . . . . .	55
3.2.4	Aplicação à Série de Furnas . . . . .	58
3.2.5	Aplicação à Série de Sobradinho . . . . .	60
3.3	Conclusão . . . . .	66
<b>4</b>	<b>Modelo Normal Generalizada Autorregressivo Periódico</b>	<b>67</b>
4.1	Modelo Normal Generalizada Autorregressivo Periódico . . . . .	68
4.2	Estimação de Máxima Verossimilhança . . . . .	69
4.3	Previsão . . . . .	71
4.4	Estudo de Simulação . . . . .	72
4.5	Aplicação à Série de Furnas . . . . .	79
4.6	Conclusão . . . . .	87
<b>5</b>	<b>Modelo Assimétrico Normal Generalizada Autorregressivo Periódico</b>	<b>88</b>
5.1	Modelo Log-Normal Generalizada Autorregressivo Periódico . . . . .	89

5.1.1	Previsão . . . . .	89
5.1.2	Aplicação à Série de Furnas . . . . .	90
5.2	Modelo Box-Cox Normal Generalizada Autorregressivo Periódico . . . . .	96
5.2.1	Estimação de Máxima Verossimilhança . . . . .	97
5.2.2	Previsão . . . . .	99
5.2.3	Estudo de Simulação . . . . .	101
5.2.4	Aplicação à Série de Furnas . . . . .	104
5.2.5	Aplicação à Série de Sobradinho . . . . .	105
5.3	Conclusão . . . . .	112
<b>6</b>	<b>Modelo Normal Generalizada Zero Inflacionada ARMA</b>	<b>114</b>
6.1	Distribuição Normal Generalizada Zero Inflacionada . . . . .	115
6.1.1	Função de Verossimilhança . . . . .	115
6.2	Modelo Normal Generalizada Zero Inflacionada ARMA . . . . .	117
6.2.1	Função de Verossimilhança . . . . .	118
6.2.2	Aplicação . . . . .	119
6.3	Conclusão . . . . .	124
<b>7</b>	<b>Conclusões e Propostas Futuras</b>	<b>126</b>
	<b>Referências</b>	<b>129</b>
	<b>Apêndice</b>	<b>133</b>
<b>A</b>	<b>Algumas Considerações sobre Processo Estocástico</b>	<b>134</b>
<b>B</b>	<b>Algumas Demonstrações</b>	<b>136</b>

# Capítulo 1

## Introdução

O estudo de séries temporais envolve observações dependentes ao longo do tempo, por exemplo, estudar o lucro de uma determinada empresa em intervalos trimestrais (Shumway e Stoffer, 2011), avaliar o crescimento ou decréscimo do preço de uma determinada ação da bolsa de valores de São Paulo no decorrer do ano (Morettin, 2008 e Morettin e Tolo, 2006) ou ainda acompanhar a disseminação de uma certa doença na população durante um determinado período de tempo (Zeger, 1988). Note que todos esses exemplos envolvem observações ordenadas no tempo, sendo que de alguma maneira valores que ocorreram em um tempo passado, não muito distante, podem influenciar os valores presentes e futuros do processo. Modelar essa dependência e utilizá-la para o cálculo de previsões de valores futuros pode trazer grandes vantagens.

Nas últimas décadas, vários modelos de séries temporais têm sido propostos na literatura. O artigo proposto por Cox (1981) categoriza modelos de séries temporais com parâmetros variando no tempo em duas classes: modelos *observation driven* e modelos *parameter driven*. Na abordagem *observation driven*, a variação dos parâmetros no tempo é introduzida fazendo os parâmetros dependerem de variáveis exógenas e de observações passadas. Embora os parâmetros sejam dependentes, eles são perfeitamente previsíveis dadas as informações passadas. Esta abordagem simplifica a avaliação da função de verossimilhança e explica o porquê que esses modelos têm se tornado populares nas mais diferentes aplicações de séries temporais. Exemplos de modelos classificados como *observation driven* são: o modelo heterocedástico condicional autorregressivo (ARCH) proposto por Engle (1982), o modelo heterocedástico condicional autorregressivo genera-

---

lizado (GARCH) introduzido por Bollerslev (1986), o modelo para dados de contagem utilizando a distribuição Poisson de Davis *et al.* (2003) e o modelo autorregressivo e de médias móveis generalizado (GARMA) proposto por Benjamin *et al.* (2003).

A alternativa para modelos *observation driven* são os modelos *parameter driven*. Nesses modelos os parâmetros são processos estocásticos que estão sujeitos à sua própria fonte de erro. Dadas as observações passadas, os parâmetros não são perfeitamente previsíveis. Como exemplos temos: o modelo para dados de contagem proposto por Zeger (1988), o modelo utilizando a distribuição binomial negativa para dados de contagem de Davis e Wu (2009), o modelo de volatilidade estocástica (SV) (ver Shephard (2005) para detalhes e discussões) e o modelo de intensidade estocástica proposto por Bauwens e Hautsch (2006). Estimação é usualmente mais difícil para esses modelos porque as funções de verossimilhança associadas não estão disponíveis em forma fechada.

O modelo ARMA proposto por Box e Jenkins (1976) é categorizado como um modelo *observation driven* e é uma abordagem padrão em estudos de séries temporais. Nessa abordagem geralmente supõe-se que os ruídos seguem uma distribuição gaussiana, porém existem séries que violam essa suposição, como por exemplo, séries cuja distribuição apresentam caudas mais pesadas ou mais leves que as da distribuição gaussiana, séries de dados de contagem, séries com distribuições assimétricas e séries com variância não constante. Isso vem motivando muitos autores a estudar processos não gaussianos, podemos citar por exemplo Rocha e Cribari-Neto (2009) que propuseram o modelo  $\beta$ ARMA para variáveis aleatórias contínuas que assumem valores no intervalo unitário padrão (0,1), Jose *et al.* (2008) que propuseram o modelo autorregressivo de primeira ordem com distribuição marginal normal-Laplace, Tomy e Jose (2009) que apresentaram o modelo autorregressivo que utiliza a distribuição normal-Laplace generalizada e Davis *et al.* (2000) que apresentaram um modelo para dados de contagem utilizando a distribuição Poisson.

O modelo GARMA, que generaliza o modelo ARMA considerando séries com distribuição da família exponencial, modelando a média da distribuição, foi proposto por Benjamin *et al.* (2003), permitindo assim adotar diferentes distribuições conforme as características da série, como por exemplo, a distribuição gama caso a série apresente apenas observações positivas, ou adotar a distribuição Poisson, binomial ou binomial negativa caso a série represente um processo de contagem.

---

Neste trabalho propomos uma nova classe de modelos de séries temporais que é a junção das ideias do modelo GARMA com a distribuição normal generalizada, introduzida por Nadarajah (2005). Adotamos essa distribuição pois as distribuições normal e Laplace são casos particulares e suas caudas podem ser mais leves ou mais pesadas do que as da distribuição normal, proporcionando, assim, uma maior flexibilidade para a modelagem de séries temporais em relação aos tradicionais modelos ARMA com distribuição normal.

Outras extensões dos modelos ARMA são os modelos periódicos autorregressivos ou ainda periódicos autorregressivos e de médias móveis, conhecidos como PAR ou PARMA, respectivamente. Os modelos propostos por Box e Jenkins (1976) consideram que a correlação é constante em relação ao tempo, o que frequentemente não é observado, por exemplo, em séries de vazões de rios. Segundo Tiao e Grupe (1980) essas séries podem ser especificadas erroneamente por meio dos modelos ARMA. Os modelos PAR são estudados em diversos trabalhos, como por exemplo, no livro de Hipel e McLeod (1994), em Pagano (1978) que estudou propriedades assintóticas, Troutman (1979) que apresentou condições para a estacionariedade e em McLeod (1993) que propôs um novo diagnóstico para a correlação periódica.

Analisamos várias séries de vazões das usinas hidrelétricas do Brasil, notamos que na maioria delas, após eliminarmos o efeito da sazonalidade, a série resultante apresenta caudas mais pesadas do que as caudas da distribuição normal. Para uma melhor modelagem dessas séries, propomos o modelo autorregressivo periódico com distribuição normal generalizada em vez dos tradicionais modelos periódicos com distribuição normal. A substituição da distribuição irá proporcionar uma maior flexibilidade no ajuste dos modelos PAR, pois a distribuição proposta é simétrica e pode apresentar curtose diferente da distribuição normal.

Em diversos estudos relacionados a hidrologia, observamos a utilização da função logarítmica para a diminuição da assimetria da distribuição das séries temporais. Em trabalhos recentes, nas mais diferentes áreas, notamos a adoção da transformação de Box-Cox, proposta em Box e Cox (1964), para esse fim, citamos, por exemplo, os trabalhos de Zheng e Song (2014), Gonçalves e Meddahi (2011) e Tsiotas (2009) que envolvem a transformação em questão com a volatilidade estocástica e o trabalho de Cordeiro e Andrade (2009) no qual propuseram um modelo que unifica a transformação de Box-Cox

com os modelos GARMA.

Propomos também neste trabalho a utilização da transformação de Box-Cox juntamente com as modificações sugeridas dos modelos ARMA e PAR. Essa junção resulta em modelos nos quais as suposições de simetria e normalidade da série não são necessárias, ou seja, os modelos propostos se credenciam para serem utilizados em séries que apresentam assimetria e curtose diferente da distribuição normal.

Como um exemplo de série temporal que não apresenta distribuição normal, temos a série da quantidade de precipitação pluviométrica diária, que assume valores reais não negativos e com excesso de zero. Para a análise de séries com essa peculiaridade, propomos uma modelagem utilizando uma modificação da distribuição normal generalizada, incorporando a inflação de zeros e levando em consideração as estruturas autorregressiva e de médias móveis.

## 1.1 Modelo ARMA

Como já foi dito anteriormente, o modelo ARMA é bastante utilizado em séries temporais. Será apresentada aqui apenas uma breve introdução, mais detalhes sobre o modelo ARMA podem ser encontrados em Box *et al.* (2008).

Considere  $\{Z_t, t \in \mathbb{Z}\}$  uma série temporal, então o modelo ARMA( $p, q$ ) é dado por

$$\begin{aligned} Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} &= a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \\ (1 - \phi_1 B - \dots - \phi_p B^p) Z_t &= (1 - \theta_1 B - \dots - \theta_q B^q) a_t \\ \Phi(B) Z_t &= \Theta(B) a_t, \end{aligned}$$

sendo que  $B$  é o operador retardo ou *backward*,  $\Phi(B)$  é o polinômio autorregressivo,  $\Theta(B)$  é o polinômio de médias móveis e  $a_t$  é um ruído branco com  $E(a_t) = 0$ ,  $Var(a_t) = \sigma^2$  e  $E(a_t a_{t-k}) = 0$  para  $k \neq 0$ .

O processo ARMA( $p, q$ ) é estacionário se as raízes de  $\Phi(B) = 0$  estiverem fora do círculo de raio unitário, e o processo é invertível se as raízes de  $\Theta(B) = 0$  estiverem fora do círculo de raio unitário (as definições de estacionariedade e invertibilidade podem ser vistas no Apêndice A).



## 1.2 Modelo GARMA

Os modelos lineares generalizados (MLG), propostos por Nelder e Wedderburn (1972), modelam a média da distribuição de interesse envolvendo três componentes, sendo essas, a componente aleatória, que é uma variável aleatória que pertence a família exponencial com média  $\mu$ , a componente sistemática, que são as variáveis explicativas que normalmente são denotadas por  $\eta = \mathbf{x}'\boldsymbol{\beta}$ , sendo que  $\mathbf{x}$  são as covariáveis observadas e  $\boldsymbol{\beta}$  é um vetor de parâmetros, e a função de ligação, aqui denotada por  $g$ , que é uma função que relaciona as médias dos componentes aleatórios com a sistemática,  $g(\mu) = \eta = \mathbf{x}'\boldsymbol{\beta}$ , para mais detalhes sobre MLG ver McCullagh e Nelder (1989).

Benjamin *et al.* (2003) introduziram os modelos GARMA (generalizado autorregressivo e de médias móveis) baseando-se nos modelos lineares generalizados (MLG) de Nelder e Wedderburn (1972), acrescentando os termos autorregressivos e de médias móveis na componente sistemática.

Seja  $\{Z_t, t \in \mathbb{Z}\}$  um processo estocástico com distribuição condicional na família exponencial, com

$$f_{Z_t}(z|\mathfrak{F}_{t-1}) = \exp \left\{ \frac{1}{\varphi} [z\vartheta_t - b(\vartheta_t)] + c(z, \varphi) \right\}, \quad (1.1)$$

sendo que  $\vartheta_t$  e  $\varphi$  representam os parâmetros canônico e de dispersão, respectivamente, as funções  $b(\cdot)$  e  $c(\cdot)$  definem uma distribuição específica da família exponencial e  $\mathfrak{F}_{t-1}$  é o conjunto de informações das observações até o instante  $t - 1$ , definido como

$$\mathfrak{F}_{t-1} = \{\mathbf{x}_1, \dots, \mathbf{x}_t, z_1, \dots, z_{t-1}, \mu_1, \dots, \mu_{t-1}\}, \quad (1.2)$$

em que  $z_i$  é o valor observado no instante  $i$ ,  $i = 1, \dots, t - 1$ ,  $\mu_i$  é a média da distribuição no instante  $i$ ,  $i = 1, \dots, t - 1$  e  $\mathbf{x}_i$  são as covariáveis observadas no instante  $i$ ,  $i = 1, \dots, t$ .

Assim como nos MLG, temos que  $E(Z_t|\mathfrak{F}_{t-1}) = \mu_t = b'(\vartheta_t)$  e  $Var(Z_t|\mathfrak{F}_{t-1}) = \varphi b''(\vartheta_t)$ . Nos modelos GARMA escrevemos  $\mu_t$  da forma

$$g(\mu_t) = \eta_t = \mathbf{x}_t'\boldsymbol{\beta} + \tau_t,$$

sendo  $\tau_t = \sum_{j=1}^p \phi_j A(z_{t-j}, \mathbf{x}_{t-j}, \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j M(z_{t-j}, \mu_{t-j})$ , sendo que  $A$  e  $M$  são funções que representam os termos autorregressivos e de médias móveis. Segundo Benjamin *et al.*

(2003), o modelo apresentado é muito geral e pode ser particularizado para uma forma mais conveniente para aplicações. Então o preditor linear pode ser escrito como

$$g(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j(g(z_{t-j}) - \mathbf{x}'_{t-j} \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j(g(z_{t-j}) - \eta_{t-j}),$$

sendo este mais flexível e parcimonioso, lembrando que  $g$  é uma função de ligação apropriada.

O modelo ARMA é um caso particular do modelo GARMA quando adotamos a distribuição normal, a função de ligação identidade, sem as variáveis exógenas e assumindo  $a_t = z_t - \mu_t$ . O modelo GARMA apresenta uma flexibilidade muito grande, pois com ele podemos ajustar modelos que envolvam séries que apresentam, por exemplo, apenas valores inteiros, adotando as distribuições Poisson, binomial ou binomial negativa, ou que apresente apenas valores reais positivos, adotando a distribuição gama ou inversa gaussiana.

### 1.3 Série Temporal Periódica

Seja  $\{x_t, t \in \mathbb{N}\}$  uma série temporal periódica com período ou núcleo de tamanho  $S$ . Faremos algumas considerações para o índice de tempo,  $t$ , em que a série é observada. Primeiramente, supomos que o número de observações da série dentro de cada período seja o mesmo e introduzimos um índice  $m$  para cada observação dentro do período, assim  $m = 1, 2, \dots, S$ . Além disso, consideramos que a série foi observada por um número inteiro de períodos  $n = NS$  e introduzimos o índice  $r$  para representar um ciclo qualquer, assim,  $r = 1, 2, \dots, N$ .

Dessa forma, o índice  $t$  referente a qualquer observação  $x_t$  pode ser escrito como função dos dois índices  $r$  e  $m$  denotado por

$$t = t(r, m) = (r - 1)S + m, \quad (1.3)$$

sendo  $r = 1, 2, \dots, N$  e  $m = 1, 2, \dots, S$ . Assim forma, para qualquer inteiro  $j \geq 0$  temos que

$$t - j = t(r, m) - j > 0, \quad (1.4)$$

para todo  $j \leq (r-1)S + m$ . No caso de dados mensais  $S = 12$ ,  $m$  representa o mês e  $r$  o ano, já para o caso de observações trimestrais,  $S = 4$ ,  $m$  refere-se ao trimestre e  $r$  ao ano.

Consideramos que o processo estocástico  $X_{t(r,m)}$  que dá origem à série temporal  $x_{t(r,m)}$  tenha as seguintes propriedades:

- i)  $E\{X_{t(r,m)}\} = \mu_{t(r,m)}$ , sendo que  $\mu_{t(r,m)}$  é uma função periódica de período  $S$ , tal que  $\mu_t = \mu_{t+KS}$  para qualquer  $K \in \mathbb{N}$ . Dessa forma  $\mu_{t(r,m)}$  fica completamente definida pelo índice  $m$ ,  $\mu_{t(r,m)} = \mu_m$ . Ou seja, supondo  $r_1, r_2 \in \mathbb{N}$  com  $r_1 < r_2$  e  $K \in \mathbb{N}$  tal que  $r_2 - r_1 = K$ , temos que

$$\begin{aligned} t(r_2, m) &= (r_2 - 1)S + m \\ &= (r_1 + K - 1)S + m \\ &= (r_1 - 1)S + m + KS \\ &= t(r_1, m) + KS, \end{aligned}$$

logo  $\mu_{t(r_2,m)} = \mu_{t(r_1,m)+KS} = \mu_{t(r_1,m)}$ . Portanto  $\mu_{t(r,m)}$  é uma função somente do índice  $m$ . Consideramos para  $\mu_m$  uma função não paramétrica dada por

$$\hat{\mu}_m = \frac{1}{N} \sum_{r=1}^N X_{t(r,m)}, m = 1, \dots, S. \quad (1.5)$$

- ii)  $Var\{X_{t(r,m)}\} = \sigma_{t(r,m)}^2$ , sendo que  $\sigma_{t(r,m)}^2$  também é uma função periódica de período  $S$ , com as mesmas propriedades da média. Temos que  $\sigma_t^2 = \sigma_{t+KS}^2$  para qualquer  $K \in \mathbb{N}$ . Desta forma  $\sigma_{t(r,m)}^2 = \sigma_m^2$ . Adotamos para  $\sigma_m^2$  uma função não paramétrica dada por

$$\hat{\sigma}_m^2 = \frac{1}{N} \sum_{r=1}^N [X_{t(r,m)} - \hat{\mu}_m]^2, m = 1, \dots, S. \quad (1.6)$$

Para simplificar a notação consideramos as funções  $\mu_m$  e  $\sigma_m^2$  para representar  $\mu_{t(r,m)}$  e  $\sigma_{t(r,m)}^2$ , respectivamente.

Quando o processo estocástico apresenta sazonalidade, um método largamente adotado para eliminar a sazonalidade (em inglês *deseasonalization*), utilizado por exemplo em Hipel e McLeod (1994) e Souza *et al.* (2012), é subtrair a sazonalidade estimada, ou seja, o processo livre de sazonalidade  $Z_{t(r,m)}$  é dado por

$$Z_{t(r,m)} = \frac{X_{t(r,m)} - \hat{\mu}_m}{\hat{\sigma}_m}, \quad (1.7)$$

com  $r \geq 1$  e  $m = 1, 2, \dots, S$ , assim  $E\{Z_{t(r,m)}\} = 0$  e  $Var\{Z_{t(r,m)}\} = 1$ . Optamos por denotar o índice de tempo como uma função de  $r$  e  $m$ ,  $t(r, m)$ , pois temos interesse em saber quem são os índices  $r$  e  $m$ . Quando não houver esse interesse, para simplificação de notação, adotaremos apenas  $t$ .

Considerando os processos estocásticos  $X_{t(r,m)}$  e  $Z_{t(r,m)}$ , com  $r \geq 1$  e  $m = 1, 2, \dots, S$ , notemos que para qualquer  $K > 1$

$$\begin{aligned}
Corr(X_{t(r,m)}, X_{t(r,m)+K}) &= \frac{E\{(X_{t(r,m)} - \mu_m)(X_{t(r,m)+K} - \mu_{m+K})\}}{\sqrt{Var\{X_{t(r,m)}\}Var\{X_{t(r,m)+K}\}}} \\
&= \frac{E\{(X_{t(r,m)} - \mu_m)(X_{t(r,m)+K} - \mu_{m+K})\}}{\sigma_m \sigma_{m+K}} \\
&= E\left\{\left[\frac{X_{t(r,m)} - \mu_m}{\sigma_m}\right]\left[\frac{X_{t(r,m)+K} - \mu_{m+K}}{\sigma_{m+K}}\right]\right\} \\
&= E(Z_{t(r,m)}Z_{t(r,m)+K}) \\
&= Corr(Z_{t(r,m)}, Z_{t(r,m)+K}), \tag{1.8}
\end{aligned}$$

portanto os processos  $X_{t(r,m)}$  e  $Z_{t(r,m)}$  possuem a mesma estrutura de correlação, ou seja, retirando o efeito sazonal por meio da equação (1.7) eliminamos o efeito periódico de  $\mu$  e  $\sigma^2$ , mas não da correlação, caso exista.

Observamos ainda que a correlação dos processos não depende do índice  $r$ , pois

$$\begin{aligned}
Corr(Z_{t(r_2,m)}, Z_{t(r_2,m)+K}) &= \frac{E\{(Z_{t(r_2,m)} - \mu_m)(Z_{t(r_2,m)+K} - \mu_{m+K})\}}{\sqrt{Var\{Z_{t(r_2,m)}\}Var\{Z_{t(r_2,m)+K}\}}} \\
&= \frac{E\{(Z_{t(r_1,m)+KS} - \mu_m)(Z_{t(r_1,m)+KS+K} - \mu_{m+K})\}}{\sigma_m \sigma_{m+K}} \\
&= \frac{E\{(Z_{t(r_1,m)} - \mu_m)(Z_{t(r_1,m)+K} - \mu_{m+K})\}}{\sigma_m \sigma_{m+K}} \\
&= Corr(Z_{t(r_1,m)}, Z_{t(r_1,m)+K}),
\end{aligned}$$

sendo  $r_2$  e  $r_1$  como definidos anteriormente.

A correlação dos processos periódicos é conhecida na literatura como correlação periódica e é definida como

$$\rho^m(K) = Corr(Z_{t(r,m)}, Z_{t(r,m)+K}) = \frac{E\{(Z_{t(r,m)} - \mu_m)(Z_{t(r,m)+K} - \mu_{m+K})\}}{\sqrt{Var\{Z_{t(r,m)}\}Var\{Z_{t(r,m)+K}\}}}, \tag{1.9}$$

sendo uma função dos parâmetros  $K$  e  $m$ , com  $m = 1, \dots, S$ .

Quando o processo estocástico não possuir periodicidade então  $S = 1$ ,  $m = 1$ ,  $t = r = 1, \dots, n$  e a correlação periódica se reduz à tradicional correlação, dependendo apenas da diferença entre os tempos.

**Proposição 1.1** *Se o processo sem sazonalidade  $Z_{t(r,m)}$ , definido na equação (1.7), é um processo  $ARMA(p,q)$ , então o processo  $X_{t(r,m)}$  é um processo periódico  $ARMA(p,q)$ .*

**Prova 1.1** *Seja  $Z_{t(r,m)}$  dado por*

$$Z_{t(r,m)} = \sum_{j=1}^p \phi_j Z_{t(r,m)-j} + a_{t(r,m)} - \sum_{j=1}^q \theta_j a_{t(r,m)-j}, \quad (1.10)$$

em que  $\{a_{t(r,m)}; r \geq 1, m = 1, 2, \dots, S\}$  é um ruído branco com

- i)  $E\{a_{t(r,m)}\} = 0$ ,
- ii)  $E\{a_{t(r,m)}a_{t(r,m)+K}\} = 0, K \neq 0$ ,
- iii)  $E\{a_{t(r,m)}^2\} = \sigma_a^2$ .

Portanto, uma vez que  $E\{Z_{t(r,m)}\} = 0$  e  $Var\{Z_{t(r,m)}\} = 1$  temos:

$$X_{t(r,m)} = \mu_{t(r,m)} + \sigma_{t(r,m)}Z_{t(r,m)}. \quad (1.11)$$

## 1.4 Proposta da Tese

A escolha de um modelo estatístico para a análise de séries temporais tem que ser feita com muito cuidado, pois o modelo escolhido tem que satisfazer as peculiaridades da série em estudo. Como exemplo, se a série apresenta apenas valores inteiros ou no intervalo  $[0, 1]$  então o ideal é que o modelo adotado também tenha essa característica. Além disso, deve-se definir, por exemplo, a utilização entre um modelo *observation driven* ou *parameter driven*.

Entre as considerações, nos preocupamos principalmente com três características, a saber:

- i) A curtose, ou seja, o quanto que as caudas da distribuição dos dados são mais leves ou mais pesadas que as da distribuição normal;
- ii) A assimetria, ou seja, quando constatamos que a distribuição dos dados reais apresenta assimetria;

iii) A correlação periódica, ou seja, quando observamos que a série é oriunda de um processo estocástico periódico.

Motivados por séries que apresentam simetria porém com curtose diferente da distribuição normal, propomos o modelo ARMA com distribuição normal generalizada (NG-ARMA), que inspirado no modelo GARMA modela a média da série. Como extensão do modelo NG-ARMA propomos o modelo ARMA com distribuição log-normal generalizada (LNG-ARMA) que se apresenta como uma alternativa para modelar séries assimétricas. Também estudamos o modelo NG-ARMA com a transformação de Box-Cox (BC-NG-ARMA), que tem como caso particular o modelo LNG-ARMA, podendo ser considerado para séries que apresentam assimetria cujo logarítmico não é a transformação mais adequada para tornar a distribuição da série simétrica.

Interessados em apresentar uma modelagem mais flexível para séries periódicas, propomos o modelo normal generalizada autorregressivo periódico (NG-PAR), que se mostra como uma alternativa para a análise de séries que exibem simetria, mas com função de autocorrelação periódica. Definimos também os modelos log-normal generalizada autorregressivo periódico (LNG-PAR) e normal generalizada autorregressivo periódico com transformação Box-Cox (BC-NG-PAR), que são úteis para a modelagem de séries periódicas assimétricas.

Determinados em analisar séries temporais que exibem uma grande quantidade de observações zeros, propomos uma reformulação da distribuição normal generalizada incorporando excesso de zeros, chamando-a de distribuição normal generalizada zero inflacionada (NGZI). A partir dela, definimos o modelo normal generalizada zero inflacionada autorregressivo e de médias móveis (NGZI-ARMA).

## 1.5 Organização do Texto

Esta tese está organizada nos seguintes capítulos:

No Capítulo 2 apresentamos o modelo NG-ARMA com algumas propriedades, com um método para o cálculo de previsões e a função de verossimilhança. Avaliamos o modelo proposto por meio de alguns estudos de simulações. Exemplificamos as aplicações

deste modelo a partir de três séries reais de diferentes áreas. Este modelo deu origem ao trabalho Milani *et al.* (2015).

No Capítulo 3 apresentamos o modelo LNG-ARMA, com uma transformação logarítmica da série original e utilizamos o modelo NG-ARMA para fazermos o cálculo das previsões. O modelo BC-NG-ARMA também é apresentado neste capítulo, juntamente com a descrição da função de verossimilhança, um método de previsão, um estudo de simulação para verificar as propriedades frequentistas do estimador de máxima verossimilhança. Duas aplicações com séries de vazões médias mensais são apresentadas.

No Capítulo 4 definimos o modelo NG-PAR, mostramos a função de verossimilhança e apresentamos um procedimento para prever valores futuros da série analisada. Ilustramos a aplicação do modelo proposto por meio de uma série de vazões médias mensais.

No Capítulo 5 formulamos os modelos LNG-PAR e BC-NG-PAR, sendo que o primeiro é um caso particular do segundo, expomos as funções de verossimilhança e métodos para o cálculo de previsões. Demonstramos a aplicação dos modelos propostos através das série de vazões médias mensais das usinas hidrelétricas de Furnas e Sobradinho. Os resultados dos modelos NG-PAR e LNG-PAR são apresentados no artigo “Generalized Normal Periodic Autoregressive Model Applied to Area of Hydrology” a ser publicado.

No Capítulo 6 definimos a distribuição normal generalizada zero inflacionada (NGZI) e a consideramos para desenvolver o modelo para séries temporais denominado de NGZI-ARMA. Expomos a função de verossimilhança condicional do modelo NGZI-ARMA. A partir da série da quantidade de precipitação pluviométrica da cidade de São Carlos exemplificamos o modelo proposto.

No Capítulo 7 apresentamos as conclusões e algumas propostas para futuras pesquisas que podem ser realizadas como continuação desta tese.

## Capítulo 2

# Modelo Normal Generalizada ARMA

A distribuição normal generalizada tem as distribuições normal e Laplace como casos particulares, foi proposta por Nadarajah (2005) e tem como uma de suas principais características o fato de que suas caudas podem ser mais leves ou mais pesadas do que as da distribuição normal.

Motivados por séries que apresentam simetria mas com curtose diferente da distribuição normal, neste capítulo apresentamos um modelo *observation driven* denominado por normal generalizada ARMA (NG-ARMA) que modela a média da série temporal da mesma forma que o modelo GARMA. Na Seção 2.1 apresentamos a distribuição normal generalizada e algumas de suas propriedades. Na Seção 2.2 mostramos o modelo NG-ARMA com algumas propriedades, um método para fazer previsões e estudos de simulação para analisar o estimador de máxima verossimilhança. Na Seção 2.3 exemplificamos três aplicações do modelo proposto ajustando-o à séries temporais das áreas de hidrologia, políticas públicas e economia. As conclusões são apresentadas na Seção 2.4.



## 2.1 Distribuição Normal Generalizada

Uma variável aleatória  $X$  tem distribuição normal generalizada com parâmetros  $\mu$ ,  $\sigma$  e  $s$  se a função densidade de probabilidade é dada por

$$f(x) = \frac{s}{2\sigma\Gamma(1/s)} \exp\left(-\left|\frac{x-\mu}{\sigma}\right|^s\right), \quad (2.1)$$

sendo que  $-\infty < x < \infty$ ,  $-\infty < \mu < \infty$ ,  $\sigma > 0$  e  $s > 0$ .

Algumas distribuições são casos particulares da distribuição normal generalizada. Para  $s = 2$  temos a distribuição normal e para  $s = 1$  temos a distribuição Laplace. Uma propriedade interessante da distribuição normal generalizada em relação à normal é a flexibilidade das caudas, ou seja, as caudas da distribuição normal generalizada podem ser mais leves ou mais pesadas do que as da distribuição normal dependendo do valor de  $s$  (Nadarajah, 2005). Algumas formas da função densidade de probabilidade são apresentadas na Figura 2.1.

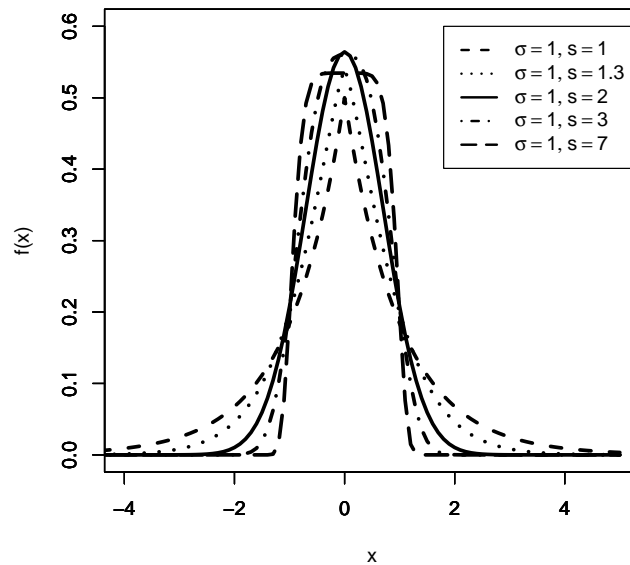


Figura 2.1: Formas da distribuição normal generalizada

O  $n$ -ésimo momento da distribuição normal generalizada é dado por

$$E(X^n) = \frac{\mu^n \sum_{k=0}^n (\sigma/\mu)^k [1 + (-1)^k] \Gamma[(k+1)/s]}{2\Gamma(1/s)}, \quad (2.2)$$

para  $n=1$ , temos  $E(X) = \mu$ . Além disso, o  $n$ -ésimo momento central é dado por

$$E[(X - \mu)^n] = \frac{\sigma^n [1 + (-1)^n] \Gamma[(n+1)/s]}{2\Gamma(1/s)}, \quad (2.3)$$

sendo que para  $n = 2$  temos que  $Var(X) = \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)}$ .

A partir da equação (2.3) observamos que o terceiro momento central é igual a zero, indicando assim que a distribuição é simétrica. A curtose, que é definida em Mood *et al.* (1974) como sendo o nível de “achatamento” da densidade perto do seu centro, para a distribuição normal generalizada é dada por

$$\text{Curtose}(X) = \frac{E[(X - \mu)^4]}{(E[(X - \mu)^2])^2} - 3 = \frac{\Gamma(5/s)\Gamma(1/s)}{(\Gamma(1/s))^2} - 3,$$

observamos que o valor da curtose depende do valor do parâmetro  $s$ . Analisamos melhor o comportamento da curtose por meio da Figura 2.2. Por ela notamos que a distribuição normal generalizada apresenta curtose maior do que a da distribuição normal para valores de  $s$  menor que 2, o contrário acontece para valores de  $s$  maior do que 2. Em outras palavras, temos que a densidade da distribuição normal generalizada exibe formato mais “pontagudo” ou mais “vago” em torno do centro da distribuição em comparação à densidade da distribuição normal, quando  $s$  assume valores abaixo ou acima de 2, respectivamente.

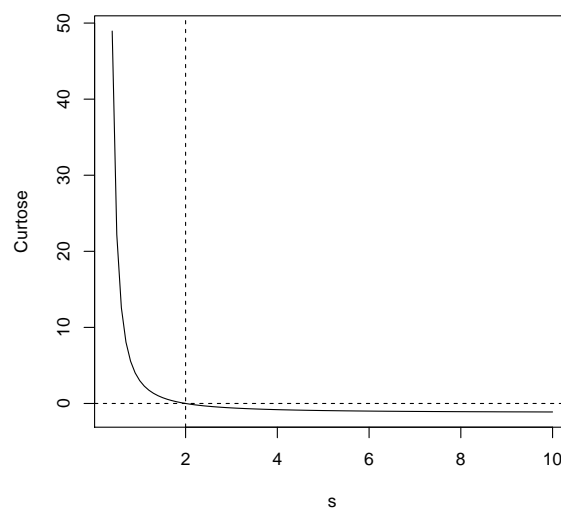


Figura 2.2: Variação da curtose versus valores de  $s$

**Proposição 2.1** *Sejam  $X_1$  e  $X_2$  variáveis aleatórias independentes, sendo  $X_1$  com distribuição uniforme contínua no intervalo  $[-1, 1]$  e  $X_2$  com distribuição gama (forma =  $1 + 1/s$ , escala = 1). Se  $X$  é uma variável aleatória com distribuição normal generalizada com os parâmetros  $\mu$ ,  $\sigma > 0$  e  $s > 0$  então  $X$  pode ser reescrita como*

$$X = \sigma X_1 X_2^{1/s} + \mu.$$

A prova da proposição 2.1 se encontra no Apêndice B.

A Proposição 2.1 é extremamente útil quando estamos interessados em simular valores da distribuição normal generalizada, pois conseguimos simular valores da distribuição normal generalizada a partir de valores simulados das distribuições uniforme e gama, uma vez que na maioria dos *softwares* estatísticos a distribuição normal generalizada não está implementada.

**Proposição 2.2** *Sejam  $X|X_2$  e  $X_2$  variáveis aleatórias, sendo  $X|X_2$  com distribuição uniforme contínua no intervalo  $[\mu - x_2^{1/s}\sigma, \mu + x_2^{1/s}\sigma]$  e  $X_2$  com distribuição gama (forma =  $1 + 1/s$ , escala = 1). Se  $X$  é uma variável aleatória com distribuição normal generalizada com os parâmetros  $\mu$ ,  $\sigma > 0$  e  $s > 0$  então a função densidade de probabilidade de  $X$  pode ser reescrita como*

$$f_X(x) = \int_0^\infty f_{X|X_2}(x|x_2) f_{X_2}(x_2) dx_2.$$

A prova da proposição 2.2 se encontra no Apêndice B.

A Proposição 2.2 apresenta uma maneira de representar a distribuição normal generalizada como uma mistura de distribuições uniformes.

## 2.2 Modelo Normal Generalizada ARMA

Considere  $\{X_{t(r,m)}; r \geq 1; m = 1, \dots, S\}$  uma série temporal com  $t(r, m) = (r - 1)S + m$ . A série livre de sazonalidade é dada por

$$Z_t = \frac{X_{t(r,m)} - \mu_m}{\sigma_m} \tag{2.4}$$

com  $\mu_m = E\{X_{t(r,m)}\}$ ,  $\sigma_m = \sqrt{Var\{X_{t(r,m)}\}}$ . Suponhamos que a distribuição condicional da série  $\{Z_t; t \in \mathbb{N}\}$  dada toda a informação passada do processo possui função densidade de probabilidade normal generalizada, ou seja,

$$Z_t | \mathfrak{F}_{t-1} \sim NG(\mu_t, \sigma, s) \quad (2.5)$$

sendo que  $E[Z_t | \mathfrak{F}_{t-1}] = \mu_t$  e  $\mathfrak{F}_{t-1}$  é o conjunto de informações das observações até o instante  $t - 1$ , definido como

$$\mathfrak{F}_{t-1} = \{\mathbf{x}_1, \dots, \mathbf{x}_t, z_1, \dots, z_{t-1}, \mu_1, \dots, \mu_{t-1}\}, \quad (2.6)$$

em que  $\mathbf{x}_i$  são as covariáveis no instante de tempo  $i$ , para  $i = 1, \dots, t$ .

Similarmente como no modelo autorregressivo e de médias móveis generalizado (Benjamin *et al.*, 2003), o nosso interesse aqui é modelar a média condicional do processo estocástico, logo  $\mu_t$  é relacionada com o preditor linear,  $\eta_t$ , por uma função de ligação,  $g$ , monótona, bijetiva e duas vezes diferenciável. A relação é expressa por

$$g(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \tau_t,$$

em que  $\tau_t$  representa os termos autorregressivos e de médias móveis, sendo definido como

$$\tau_t = \sum_{j=1}^p \phi_j \{g(z_{t-j}) - \mathbf{x}'_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \theta_j \{g(z_{t-j}) - \eta_{t-j}\}. \quad (2.7)$$

Adotamos a função identidade como função de ligação, logo

$$\begin{aligned} \eta_t = \mu_t &= \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j (z_{t-j} - \mathbf{x}'_{t-j} \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) \\ &= \mathbf{x}'_t \boldsymbol{\beta} - \sum_{j=1}^p \phi_j \mathbf{x}'_{t-j} \boldsymbol{\beta} + \sum_{j=1}^p \phi_j z_{t-j} + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}), \\ &= \sum_{j=0}^p \mathbf{x}'_{t-j} \boldsymbol{\nu}_j + \sum_{j=1}^p \phi_j z_{t-j} + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}), \end{aligned} \quad (2.8)$$

sendo que  $\boldsymbol{\nu}_0 = \boldsymbol{\beta}$ ,  $\boldsymbol{\nu}_i = -\phi_i \boldsymbol{\beta}$ , para  $i = 1, \dots, p$ . Note que se utilizarmos o parâmetro  $s = 2$  e adotarmos  $z_t - \mu_t = a_t$ , em que  $a_t$  é o ruído branco, para qualquer instante de tempo  $t$ , o modelo NG-ARMA torna-se o modelo ARMAX( $p, q, p$ ) (Caines, 1988).

Caso não haja covariáveis, ou seja, assumindo  $\mathbf{x}_t = 1$ , para qualquer instante de tempo  $t$ , temos que

$$\eta_t = \mu_t = \beta_0 + \sum_{j=1}^p \phi_j z_{t-j} + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}), \quad (2.9)$$

sendo que  $\beta_0 = (1 - \sum_{j=1}^p \phi_j)\beta$ . Se utilizarmos o parâmetro  $s = 2$  e adotarmos  $z_t - \mu_t = a_t$ , o modelo NG-ARMA se transforma no modelo ARMA( $p, q$ ).

O modelo NG-ARMA que é definido pelo par de equações (2.5) e (2.9) ou por (2.5) e (2.8), apresenta uma maior flexibilidade em relação ao modelo ARMA ou ao ARMAX, respectivamente, por assumir a distribuição condicional normal generalizada em vez da distribuição normal. Portanto o modelo NG-ARMA é uma alternativa para modelar séries temporais que apresentam simetria e curtose diferente da distribuição normal.

A partir do modelo NG-ARMA a variância condicional é dada por

$$\text{Var}[Z_t | \mathfrak{F}_{t-1}] = \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)}. \quad (2.10)$$

Sejam os polinômios  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ,  $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  e  $\Psi(B) = \Phi(B)^{-1} \Theta(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$ , como definidos no Apêndice B.

**Teorema 2.1** *A esperança marginal (ou incondicional) de  $Z_t$  do modelo NG-ARMA definido pelas equações (2.5) e (2.9) ou por (2.5) e (2.8) é dada respectivamente por*

$$E[Z_t] = \frac{\beta_0}{1 - \sum_{j=1}^p \phi_j} = \beta, \quad (2.11)$$

$$E[Z_t] = \mathbf{x}'_t \boldsymbol{\beta}, \quad (2.12)$$

desde que  $\sum_{i=0}^{+\infty} \psi_i^2 < +\infty$ .

A prova do Teorema 2.1 se encontra no Apêndice B.

**Teorema 2.2** *A variância marginal (ou incondicional) de  $Z_t$  do modelo NG-ARMA é dada por*

$$\text{Var}(Z_t) = \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)} \sum_{i=0}^{\infty} \psi_i^2, \quad (2.13)$$

desde que  $\sum_{i=0}^{+\infty} \psi_i^2 < +\infty$ . Observe que a variância de  $Z_t$  é a mesma com ou sem a presença de covariáveis.

A prova do Teorema 2.2 se encontra no Apêndice B.

A variância marginal no caso  $p = q = 1$  é

$$\text{Var}(Z_t) = \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)} \left( 1 + \frac{\phi_1 + \theta_1}{1 - \phi_1} \right).$$

Considere  $\{X_{t(r,m)}; r \geq 1; m = 1, \dots, S\}$  o processo sazonal, a esperança condicional é dada por

$$E[X_{t(r,m)} | \mathfrak{F}_{t-1}] = E[\mu_m + \sigma_m Z_t | \mathfrak{F}_{t-1}] = \mu_m + \sigma_m E[Z_t | \mathfrak{F}_{t-1}] = \mu_m + \sigma_m \mu_t, \quad (2.14)$$

enquanto que a variância condicional do processo sazonal é dada por

$$\text{Var}[X_{t(r,m)} | \mathfrak{F}_{t-1}] = \text{Var}[\mu_m + \sigma_m Z_t | \mathfrak{F}_{t-1}] = \sigma_m^2 \text{Var}[Z_t | \mathfrak{F}_{t-1}] = \sigma_m^2 \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)}, \quad (2.15)$$

já a esperança marginal (ou incondicional) do processo sazonal é dado por

$$E[X_{t(r,m)}] = E[\mu_m + \sigma_m Z_t] = \mu_m + \sigma_m (\mathbf{x}'_t \boldsymbol{\beta}), \quad (2.16)$$

quando adotamos covariáveis, e

$$E[X_{t(r,m)}] = \mu_m + \sigma_m \frac{\beta_0}{1 - \sum_{j=1}^p \phi_j}, \quad (2.17)$$

quando assumimos que não haja covariáveis ( $\mathbf{x}'_t \boldsymbol{\beta} = \beta$ ). E por fim, a variância marginal (ou incondicional) do processo sazonal é dada por

$$\text{Var}[X_{t(r,m)}] = \text{Var}[\mu_m + \sigma_m Z_t] = \sigma_m^2 \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)} \sum_{i=0}^{\infty} \psi_i^2, \quad (2.18)$$

desde que  $\sum_{i=0}^{+\infty} \psi_i^2 < +\infty$ .

Além da formulação do modelo NG-ARMA apresentada anteriormente para melhor estudar séries que apresentam sazonalidade, podemos formular o modelo NG-ARMA para contemplar séries que apresentam tendência ou componente sazonal na função de autocorrelação. A seguir apresentamos estas duas novas formulações do modelo NG-ARMA.

Seja  $Z_t$  dado pela equação (2.4) e  $\{Y_t; t = 1, \dots, n\}$  a série temporal resultante da operação

$$Y_t = Z_{t+1} - Z_t, \quad (2.19)$$

ou seja, da primeira diferença da série temporal  $Z_t$ , adotamos que  $Y_t | \mathfrak{F}_{t-1} \sim NG(\mu_t, \sigma, s)$ , sendo que  $\mu_t$  é dado como na equação (2.9). Assim o modelo resultante considera que a série temporal apresenta sazonalidade, que pode ser eliminada pela equação (2.4), e uma diferença, que pode ser útil para séries não estacionárias ou com tendência.

Considere agora  $Z_t$  dado pela equação (2.4) com  $Z_t | \mathfrak{F}_{t-1} \sim NG(\mu_t, \sigma, s)$ . Mas a série resultante apresenta comportamento sazonal na função de autocorrelação, ou seja, a função de autocorrelação apresenta lags significativos com defasagem de um, dois ou mais anos. Para melhor explicar séries com estas características assumimos que  $\mu_t$  é dado por

$$\mu_t = \beta_0 + \sum_{j=1}^p \phi_j z_{t-j} + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) + \sum_{j=1}^{p^*} \phi_j^* z_{t-12j} + \sum_{j=1}^{q^*} \theta_j^* (z_{t-12j} - \mu_{t-j-11}), \quad (2.20)$$

o modelo definido é chamado de NG-ARMA com componente sazonal e será representado como NG-ARMA( $p, q$ )S( $p^*, q^*$ ).

### 2.2.1 Estimação de Máxima Verossimilhança

Denote os parâmetros do modelo para serem estimados por  $(\beta_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s)$ , sendo que  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$  e  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$ . Seja  $\mathbf{z} = \{z_1, \dots, z_n\}$  a série temporal observada,  $\mu_t$  como descrito na equação (2.9) e  $r^* = \max(p, q)$ . Considerando a função de verossimilhança condicional introduzida por Cox (1975), que é baseada inteiramente na distribuição condicional da resposta atual, dados seus valores passados, a função de verossimilhança condicionada nas  $r^*$ -primeiras observações é dada por

$$\begin{aligned} L(\beta_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \mathbf{z}) &= \prod_{t=r^*+1}^n f(z_t | \mathfrak{F}_{t-1}) \\ &= \prod_{t=r^*+1}^n \frac{s}{2\sigma\Gamma(1/s)} \exp \left\{ - \left| \frac{z_t - \mu_t}{\sigma} \right|^s \right\} \\ &= \left( \frac{s}{2\sigma\Gamma(1/s)} \right)^{n-r^*} \times \\ &\quad \exp \left\{ - \frac{1}{\sigma^s} \sum_{t=r^*+1}^n \left| z_t - \beta_0 - \sum_{j=1}^p \phi_j z_{t-j} - \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) \right|^s \right\}. \end{aligned} \quad (2.21)$$

Observamos pela equação (2.21) que a função de verossimilhança é bastante com-

plexa, e em muitos casos é conveniente trabalhar com a função log-verossimilhança, que é dada por

$$\begin{aligned} l(\beta_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \mathbf{z}) &= \log[L(\beta_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \mathbf{z})] \\ &= (n - r^*) \log \left( \frac{s}{2\sigma\Gamma(1/s)} \right) \\ &\quad - \frac{1}{\sigma^s} \sum_{t=r^*+1}^n \left| z_t - \beta_0 - \sum_{j=1}^p \phi_j z_{t-j} - \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) \right|^s. \end{aligned}$$

Devido à complexidade da equação log-verossimilhança e das suas derivadas em relação aos parâmetros, não conseguimos encontrar os estimadores de máxima verossimilhança de forma fechada. Portanto as estimativas de máxima verossimilhança são encontradas utilizando métodos iterativos, como por exemplo, o método “L-BFGS-B” (Byrd *et al.*, 1995) que está disponível no *software* R Core Team (2014). Os intervalos de confiança são construídos considerando normalidade assintótica.

## 2.2.2 Previsão

Estamos interessados em prever um valor de  $X_{t+h}$  com  $h \geq 1$ , supondo que temos as observações  $\{x_1, x_2, \dots, x_t\}$  até o instante  $t$ , que é chamando *origem* das previsões. Adotamos  $\mu_t$  como descrito na equação (2.9). O valor  $\hat{X}_t(h) = \hat{E}(X_{t+h} | \mathfrak{F}_t)$  será a previsão de *origem*  $t$  e *horizonte*  $h$ , mas para que conseguíssemos encontrar as previsões para a série original precisamos inicialmente calcular as previsões para a série  $Z_t$ , ou seja, prever um valor de  $Z_{t+h}$ , com  $h \geq 1$ . A partir das observações da série original conseguimos obter a série  $\{z_1, z_2, \dots, z_t\}$ , até o instante  $t$ , a previsão para o modelo NG-ARMA é feita através do preditor linear. Para  $h = 1$  temos a previsão a um passo a frente dada por

$$\hat{Z}_t(1) = \hat{\beta}_0 + \sum_{j=1}^p \hat{\phi}_j z_{t+1-j} + \sum_{j=1}^q \hat{\theta}_j (z_{t+1-j} - \hat{\mu}_{t+1-j}), \quad (2.22)$$

sendo que  $\hat{\beta}_0$ ,  $\hat{\phi}_j$ , e  $\hat{\theta}_i$ , para  $j = 1, \dots, p$  e  $i = 1, \dots, q$ , são as estimativas de máxima verossimilhança dos respectivos parâmetros e  $\hat{\mu}_{t+1-j}$  é encontrado adotando  $\hat{\mu}_1 = \dots = \hat{\mu}_q = 0$  e substituindo as estimativas dos parâmetros na equação (2.9). Observando a equação (2.22) notamos que  $\hat{\mu}_{t+1}$  coincide com  $\hat{Z}_t(1)$ .



Nas previsões para  $h > 1$  utilizamos o fato que

$$\widehat{E}[Z_{t+h-j}|\mathfrak{F}_t] = \begin{cases} z_{t+h-j} & \text{se } h \leq j \\ \widehat{Z}_t(h-j) & \text{se } h > j, \end{cases} \quad (2.23)$$

$$\widehat{\mu}_{t+h-j} = \begin{cases} \widehat{\mu}_{t+h-j} & \text{se } h \leq j \\ \widehat{Z}_t(h-j) & \text{se } h > j, \end{cases} \quad (2.24)$$

com isso a previsão é dada por

$$\widehat{Z}_t(h) = \widehat{\beta}_0 + \sum_{j=1}^p \widehat{\phi}_j \widehat{E}[Z_{t+h-j}|\mathfrak{F}_t] + \sum_{j=1}^q \widehat{\theta}_j (\widehat{E}[Z_{t+h-j}|\mathfrak{F}_t] - \widehat{\mu}_{t+h-j}). \quad (2.25)$$

Notamos também que quanto maior o *horizonte* de predição maior a quantidade de erros que cometemos, pois sempre assumimos que o valor que observaremos é igual à média condicional naquele instante e isso dificilmente acontece.

Para um grande *horizonte* de previsão os valores preditos convergem para a média incondicional da série. Uma sugestão quando o interesse é esse caso, é a atualização dos dados periodicamente, se possível. Por exemplo, se pretendemos fazer previsões para 30 dias e a cada dia que se passa observamos o evento de interesse, então podemos fazer um novo ajuste com esse dia a mais, e assim, fazer previsões para os outros 29 dias restantes. Repetindo esse procedimento, dias próximos do último dia observado irão apresentar erros de previsões menores quando comparados com as previsões feitas quando ajustamos a série apenas uma vez.

Observemos agora que

$$\begin{aligned} \widehat{X}_t(h) = \widehat{E}(X_{t+h}|\mathfrak{F}_t) &= \widehat{E}(\widehat{\mu}_{m_0} + \widehat{\sigma}_{m_0} Z_{t+h}|\mathfrak{F}_t) \\ &= \widehat{\mu}_{m_0} + \widehat{\sigma}_{m_0} \widehat{E}(Z_{t+h}|\mathfrak{F}_t) \\ &= \widehat{\mu}_{m_0} + \widehat{\sigma}_{m_0} \widehat{\mu}_{t+h}, \end{aligned} \quad (2.26)$$

considerando que  $t+h = (r_0-1)S+m_0$ . Portanto, a partir da equação (2.26) conseguimos prever a série original a partir das predições da série  $Z_t$ , para qualquer *horizonte*  $h \geq 1$ .

### 2.2.3 Estudo de Simulação

Para o estudo de simulação a geração de uma série temporal do modelo NG-ARMA é feita seguindo os seguintes passos:

1. definir os valores dos parâmetros  $\beta_0$ ,  $\phi$ ,  $\theta$ ,  $\sigma$  e  $s$ ;
2. definir os valores iniciais para  $\mu_t$  e  $z_t$ ;
3. calcular  $g(\mu_t) = \beta_0 + \sum_{j=1}^p \phi_j g(z_{t-j}) + \sum_{j=1}^q \theta_j (g(z_{t-j}) - \eta_{t-j})$  e gerar  $z_t$  da distribuição  $\text{NG}(\mu_t, \sigma, s)$ , para  $t = \max(p, q) + 1, \dots$ ;
4. repetir o item 3 até conseguir uma série com  $K$  valores a mais do que se deseja;
5. desconsiderar os  $K$  primeiros valores da série para conseguir o tamanho da série desejada.

A justificativa de se construir uma série com  $K$  valores a mais e depois desconsiderar os  $K$  primeiros valores é para que os valores iniciais afetem o mínimo possível a série.

A fim de verificar as propriedades empíricas dos estimadores de máxima verossimilhança, repetimos 1000 vezes o processo de geração e estimação, para diferentes tamanhos de séries ( $n=100, 200, 500$  e  $1000$ ), sendo que para cada repetição obtemos o valor das estimativas pontuais e a matriz de informação observada de Fisher. A partir destes valores, calculamos a média das estimativas (ME), o vício absoluto médio (VAM), que é dado, por exemplo para o parâmetro  $\beta_0$ , por

$$\text{VAM}(\hat{\beta}_0) = \frac{\sum_{j=1}^M |\beta_0 - \hat{\beta}_0^j|}{M}, \quad (2.27)$$

sendo que  $M$  é a quantidade de repetições,  $\beta_0$  é o verdadeiro valor do parâmetro e  $\hat{\beta}_0^j$  é a estimativa de máxima verossimilhança de  $\beta_0$  na  $j$ -ésima repetição. O erro quadrático médio (EQM), por sua vez é dado, também por exemplo para  $\beta_0$ , por

$$\text{EQM}(\hat{\beta}_0) = \frac{\sum_{j=1}^M (\beta_0 - \hat{\beta}_0^j)^2}{M - 1}, \quad (2.28)$$

e o percentual de cobertura (PC) dos parâmetros, que é o quociente entre o número de intervalos de confiança que contêm o verdadeiro valor do parâmetro e o número total de intervalos de confiança construídos, sendo que o intervalo de confiança de 95% do vetor de parâmetros é construído assumindo normalidade assintótica.

Para o estudo consideramos o modelo NG-ARMA(1,1) com os valores  $\beta_0 = 0, 1$ ;  $\phi_1 = 0, 2$ ;  $\theta_1 = 0, 35$ ;  $\sigma = 2$  e com o objetivo de investigar a influência do valor de  $s$  nas medidas de interesse adotamos  $s = 1, 5$ ;  $3, 0$  e  $4, 5$ . Os resultados estão apresentados na Tabela 2.1.

Tabela 2.1: Estudo de simulação do modelo NG-ARMA considerando as métricas VAM, ME, EQM e PC

	$s = 1, 5$				$s = 3, 0$				$s = 4, 5$			
	VAM	ME	EQM	PC	VAM	ME	EQM	PC	VAM	ME	EQM	PC
n=100												
$\beta_0$	0,1896	0,1036	0,0583	0,8750	0,1379	0,1052	0,0309	0,9320	0,1107	0,1029	0,0204	0,9360
$\phi_1$	0,1582	0,1765	0,0422	0,8450	0,1548	0,1696	0,0391	0,9110	0,1391	0,1754	0,0325	0,8970
$\theta_1$	0,1573	0,3598	0,0413	0,8130	0,1487	0,3706	0,0375	0,8780	0,1337	0,3652	0,0305	0,8780
$\sigma$	0,3128	1,9827	0,1628	0,9170	0,1642	1,9853	0,0438	0,9160	0,1175	1,9735	0,0226	0,9330
$s$	0,3276	1,6216	0,2541	0,9300	1,0910	3,6829	3,3141	0,9570	1,8578	5,7015	8,1732	0,9610
n=200												
$\beta_0$	0,1309	0,1031	0,0279	0,9310	0,0934	0,0992	0,0128	0,9450	0,0748	0,1035	0,0090	0,9390
$\phi_1$	0,1047	0,1888	0,0181	0,9190	0,1073	0,1845	0,0185	0,9350	0,0939	0,1848	0,0140	0,9240
$\theta_1$	0,1017	0,3584	0,0167	0,9230	0,1043	0,3597	0,0171	0,9300	0,0909	0,3594	0,0129	0,9110
$\sigma$	0,2114	1,9771	0,0699	0,9440	0,1083	1,9920	0,0190	0,9370	0,0795	1,9957	0,0102	0,9330
$s$	0,1905	1,5338	0,0627	0,9510	0,5500	3,2591	0,6059	0,9560	1,1963	5,2253	3,3017	0,9570
n=500												
$\beta_0$	0,0831	0,0986	0,0110	0,9490	0,0563	0,1023	0,0050	0,9480	0,0458	0,1021	0,0033	0,9530
$\phi_1$	0,0678	0,1921	0,0075	0,9290	0,0664	0,1923	0,0068	0,9430	0,0565	0,1931	0,0051	0,9390
$\theta_1$	0,0639	0,3557	0,0066	0,9370	0,0604	0,3530	0,0057	0,9490	0,0545	0,3529	0,0047	0,9470
$\sigma$	0,1331	1,9940	0,0277	0,9380	0,0655	1,9991	0,0069	0,9410	0,0478	1,9991	0,0036	0,9460
$s$	0,1175	1,5129	0,0219	0,9410	0,2954	3,0859	0,1456	0,9630	0,5374	4,7123	0,5245	0,9630
n=1000												
$\beta_0$	0,0584	0,1025	0,0054	0,9390	0,0402	0,1017	0,0025	0,9520	0,0316	0,1017	0,0016	0,9490
$\phi_1$	0,0481	0,1961	0,0036	0,9460	0,0465	0,1964	0,0034	0,9300	0,0384	0,1968	0,0024	0,9470
$\theta_1$	0,0453	0,3529	0,0032	0,9470	0,0436	0,3508	0,0030	0,9420	0,0370	0,3508	0,0022	0,9450
$\sigma$	0,0946	1,9975	0,0141	0,9390	0,0468	1,9979	0,0034	0,9480	0,0342	1,9989	0,0018	0,9580
$s$	0,0813	1,5052	0,0105	0,9450	0,1938	3,0304	0,0626	0,9530	0,3473	4,5817	0,1982	0,9600

Notamos que com o aumento do tamanho da amostra, o percentual de cobertura fica próximo do nominal, a média das estimativas se aproximam do verdadeiro valor e o vício absoluto médio e o erro quadrático médio diminuem.

Comparando as métricas estudadas para os diferentes valores de  $s$ , observamos que para todos os tamanhos de séries e para todos os parâmetros, exceto o parâmetro  $s$ , o valor das métricas EQM e VAM tendem a diminuir conforme o valor de  $s$  aumenta. Enquanto isso as métricas ME e PC parecem não ser afetadas pelo valor de  $s$ .

Com o objetivo de verificar quantos termos autorregressivos e de médias móveis devem ser incorporados no modelo para melhor explicar a série, apresentamos um estudo de simulação que segue os seguintes passos:

1. gerar uma série de tamanho  $n$  a partir de uma determinada ordem do modelo;

2. obter o valor de máxima verossimilhança e a inversa da matriz de informação observada de Fisher dos modelos com as ordens (1,1), (2,1), (1,2) e (2,2), utilizando o processo de estimação descrito anteriormente;
3. construir os intervalos de confiança dos parâmetros assumindo normalidade assintótica, para cada ordem do modelo;
4. verificar nos modelos de ordem diferente da (1,1) se os parâmetros autorregressivos e/ou de médias móveis incluídos a mais em relação a ordem (1,1) são significativos, para os modelos que não são significativos descartá-los;
5. escolher a melhor ordem segundo os critérios AIC e BIC dos modelos que não foram descartados no item anterior.
6. repetir 5000 vezes os passos de 1 até 5.

As ordens utilizadas no estudo foram as (1,1), (2,1), (1,2) e (2,2). Com os passos descritos acima temos que para cada ordem e tamanho de série escolhidos no item 1, obtemos a quantidade de vezes que cada uma das quatro ordens adotadas no estudo foram indicadas como a melhor ordem para explicar as séries simuladas, isso para cada um dos critérios AIC e BIC. A quantidade de indicações, em porcentagem, de cada critério é apresentada na Tabela 2.2.

Observamos que para as diferentes ordens utilizadas na geração, a quantidade de acertos aumenta com o aumento do tamanho da série. Observamos também que mesmo para séries de tamanho pequeno ( $n=100$ ) o critério AIC acertou em torno de 70% e o critério BIC acertou por volta de 75%, nas ordens (1,1) e (2,1). Para as ordens (1,2) e (2,2) o critério AIC acertou por volta de 50%, enquanto o critério BIC teve um desempenho um pouco pior, acertando em torno de 40%.

Para séries com  $n$  igual a 1000, notamos que os critérios acertam mais de 80%, exceto o critério AIC na ordem  $p = q = 1$ . De uma maneira geral, o critério BIC é o mais indicado, pois quando geramos das ordens (1,1) e (2,1) ele acertou mais que o critério AIC, para todos os tamanhos de série em estudo e, nas demais ordens, com o aumento do tamanho da série a quantidade de acerto ou ficou muito próxima ou até mesmo superou o desempenho do critério AIC.

Tabela 2.2: Porcentagem das indicativas da melhor ordem segundo os critérios AIC/BIC quando a geração é feita a partir do modelo NG-ARMA

Gerado	n	Estimado			
		NG-ARMA(1,1)	NG-ARMA(1,2)	NG-ARMA(2,1)	NG-ARMA(2,2)
NG-ARMA(1,1)	100	67,76/75,92	6,24/6,40	20,50/16,08	5,50/1,60
$\beta_0 = 0, 1; \phi_1 = 0, 9; \theta_1 = 0, 2;$	200	72,16/83,20	4,88/4,92	15,56/10,98	7,40/0,90
$\sigma = 2; s = 3$	500	76,08/89,82	4,22/4,06	9,20/5,62	10,50/0,50
	1000	76,86/92,24	3,74/3,24	6,88/4,20	12,52/0,32
NG-ARMA(1,2)	100	29,22/37,88	45,84/40,46	18,22/18,78	6,72/2,88
$\beta_0 = 0, 1; \phi_1 = 0, 5; \theta_1 = 0, 45;$	200	20,12/30,04	58,40/51,58	16,36/17,06	5,12/1,32
$\theta_2 = -0, 75; \sigma = 2; s = 3$	500	5,60/10,66	74,50/72,44	15,02/16,10	4,88/0,80
	1000	0,86/1,78	85,04/87,76	8,86/9,90	5,24/0,56
NG-ARMA(2,1)	100	4,10/6,32	5,60/5,98	71,90/77,90	18,40/9,80
$\beta_0 = 0, 1; \phi_1 = 0, 9; \phi_2 = -0, 4;$	200	0,20/0,84	2,76/3,06	81,40/88,58	15,64/7,52
$\theta_1 = 0, 6; \sigma = 2; s = 3$	500	0,00/0,00	0,18/0,30	88,06/94,74	11,76/4,96
	1000	0,00/0,00	0,00/0,00	90,20/97,12	9,80/2,88
NG-ARMA(2,2)	100	0,20/0,70	0,40/0,44	43,58/54,98	55,82/43,88
$\beta_0 = 0, 1; \phi_1 = 0, 95; \theta_1 = -0, 65;$	200	0,00/0,00	0,00/0,20	25,04/38,06	74,96/61,92
$\theta_1 = 0, 8; \theta_2 = -0, 35; \sigma = 2; s = 3$	500	0,00/0,00	0,00/0,00	3,26/8,76	96,74/91,24
	1000	0,00/0,00	0,00/0,00	0,08/0,54	99,92/99,46

Notemos que o critério AIC não escolheu de maneira generalizada os modelos com maiores ordens. De maneira análoga, o critério BIC não escolheu de forma sistemática os modelos com menores ordens.

Com o objetivo de verificar qual o erro que cometemos na previsão de valores futuros, geramos 1000 séries do modelo NG-ARMA(1,1) (com os parâmetros  $\beta_0 = 2, 0; \phi_1 = 0, 8; \theta = 0, 35; \sigma = 1; s = 1, 5$ ) com tamanhos iguais a 110, 210, 510 e 1010. Utilizando o método de estimação já descrito, encontramos as estimativas de máxima verossimilhança para os parâmetros do modelo NG-ARMA(1,1), sendo que os últimos dez valores de cada série não foram utilizados na estimação.

Através do método de previsão descrito na Seção 2.2.2, fizemos previsões de dez passos a frente ( $h=10$ ) e comparamos as previsões com os verdadeiros valores através dos

critérios: erro absoluto médio (EAM), que é dado por

$$\text{EAM} = \frac{\sum_{t=n+1}^{n+h} |z_t - \hat{\mu}_t|}{h}, \quad (2.29)$$

erro quadrado médio (EQM), que é dado por

$$\text{EQM} = \frac{\sum_{t=n+1}^{n+h} (z_t - \hat{\mu}_t)^2}{h}, \quad (2.30)$$

e erro médio absoluto percentual (MAPE), que é dado por

$$\text{MAPE} = 100 \frac{\sum_{t=n+1}^{n+h} \frac{|z_t - \hat{\mu}_t|}{z_t}}{h} \%, \quad (2.31)$$

sendo que  $n$  é o tamanho da série utilizada na otimização e  $\hat{\mu}_t$ 's são os valores previstos (detalhes destes e de mais métodos são apresentados em Flores (1986)). Apresentamos na Tabela 2.3 as médias e os intervalos de confiança com 95% dos valores dos critérios, sendo estes construídos a partir dos percentis 2,5% e 97,5%.

Notamos pela Tabela 2.3 que com o aumento do tamanho da série as médias dos critérios apresentaram uma pequena variação dos valores. Observamos também que a amplitude dos intervalos de confiança dos critérios tendem a diminuir com o aumento do tamanho da série.

Tabela 2.3: Avaliação das previsões feitas a partir do modelo NG-ARMA segundo os critérios MAPE, EQM e EAM

Tamanho	MAPE(%)		EQM		EAM	
	Média	IC	Média	IC	Média	IC
100	14,122	(4,314; 37,557)	2,952	(0,298; 11,153)	1,335	(0,446; 3,012)
200	13,737	(4,693; 36,143)	2,841	(0,335; 11,183)	1,301	(0,485; 2,997)
500	13,817	(4,505; 34,586)	2,691	(0,333; 9,391)	1,278	(0,468; 2,790)
1000	13,828	(4,854; 36,325)	2,736	(0,390; 10,250)	1,289	(0,498; 2,807)

Este estudo de simulação nos informa como os dez valores preditos se comportam. Para que possamos ter a sensibilidade do que acontece em cada *horizonte* repetimos o estudo anterior, mas agora calculamos as métricas para cada *horizonte* de previsão, o resultado é apresentado na Tabela 2.4.

Observamos através da Tabela 2.4 que o valor médio do MAPE cresce conforme o *horizonte* de predição aumenta, por exemplo, para o *horizonte* igual a 4 o valor do critério é o dobro do observado para o *horizonte* igual a 1. Notamos também que com o aumento do *horizonte* o valor médio do critério se estabiliza em torno de 15%.

Os valores médios dos critérios EQM e EAM também crescem com a expansão do *horizonte*, mas para o critério EQM o dobro do registrado para o *horizonte* igual a 1 já é visto para o *horizonte* igual a 2, enquanto para o critério EAM esse mesmo aumento só é visto no *horizonte* igual a 3. Observamos que para os maiores *horizontes* em estudo os valores médios dos critérios ficaram por volta de 3,5 e 1,5, respectivamente para os critérios EQM e EAM.

Tabela 2.4: Avaliação das previsões feitas a partir do modelo NG-ARMA segundo os critérios MAPE, EQM e EAM para cada *horizonte* de predição

Tamanho	MAPE		EQM		EAM	
	Média	IC	Média	IC	Média	IC
h=1						
100	7,056	(0,263; 22,161)	0,763	(0,001; 4,374)	0,671	(0,031; 2,092)
200	6,938	(0,230; 24,450)	0,715	(0,001; 4,354)	0,640	(0,023; 2,087)
500	7,400	(0,184; 23,843)	0,822	(0,001; 4,226)	0,697	(0,020; 2,056)
1000	7,027	(0,297; 22,419)	0,749	(0,001; 3,981)	0,673	(0,029; 1,995)
h=2						
100	10,832	(0,395; 33,582)	1,736	(0,002; 8,840)	1,034	(0,040; 2,973)
200	11,300	(0,294; 39,583)	1,801	(0,001; 9,736)	1,022	(0,030; 3,120)
500	11,639	(0,499; 40,022)	1,888	(0,003; 10,092)	1,077	(0,051; 3,177)
1000	10,578	(0,456; 34,744)	1,642	(0,002; 8,675)	1,003	(0,046; 2,945)
h=3						
100	13,263	(0,435; 44,792)	2,532	(0,002; 13,547)	1,257	(0,040; 3,681)
200	13,365	(0,332; 56,210)	2,533	(0,001; 13,763)	1,228	(0,034; 3,710)
500	13,265	(0,402; 44,525)	2,413	(0,002; 11,908)	1,224	(0,041; 3,451)
1000	12,916	(0,529; 41,945)	2,376	(0,003; 13,376)	1,221	(0,053; 3,657)
h=4						
100	14,577	(0,516; 47,900)	3,051	(0,003; 16,011)	1,389	(0,052; 4,001)
200	14,466	(0,545; 55,065)	3,012	(0,002; 17,096)	1,366	(0,050; 4,135)
500	14,211	(0,507; 47,131)	2,783	(0,002; 13,783)	1,312	(0,048; 3,713)
1000	13,829	(0,498; 46,968)	2,771	(0,002; 13,716)	1,310	(0,048; 3,703)
h=5						
100	15,210	(0,737; 47,670)	3,353	(0,005; 16,389)	1,461	(0,071; 4,048)
200	15,142	(0,581; 49,758)	3,310	(0,004; 17,007)	1,448	(0,064; 4,124)
500	14,721	(0,512; 48,896)	2,960	(0,003; 15,414)	1,370	(0,054; 3,926)
1000	14,917	(0,520; 53,547)	3,094	(0,003; 18,050)	1,400	(0,051; 4,249)
h=10						
100	16,984	(0,586; 54,759)	3,704	(0,003; 17,274)	1,546	(0,052; 4,156)
200	15,124	(0,794; 49,834)	3,368	(0,006; 17,549)	1,459	(0,078; 4,189)
500	15,428	(0,531; 60,891)	3,288	(0,003; 17,255)	1,407	(0,051; 4,154)
1000	16,258	(0,656; 59,909)	3,399	(0,004; 17,845)	1,456	(0,064; 4,224)

## 2.3 Aplicações

Nesta Seção analisamos o ajuste do modelo NG-ARMA em três séries temporais reais. Os dados reais são das áreas de hidrologia, políticas públicas e economia.

### 2.3.1 Série de Vazões da Usina Hidrelétrica de Furnas

Consideramos a série de vazões médias mensais afluente ( $m^3/s$ ) da usina hidrelétrica de Furnas observada de janeiro de 1931 até dezembro de 2012, apresentada na Figura 2.3(a). Devido ao fato que mais a frente faremos previsões, desconsideraremos o ano de 2012 para que posteriormente possamos comparar as previsões com as observações.

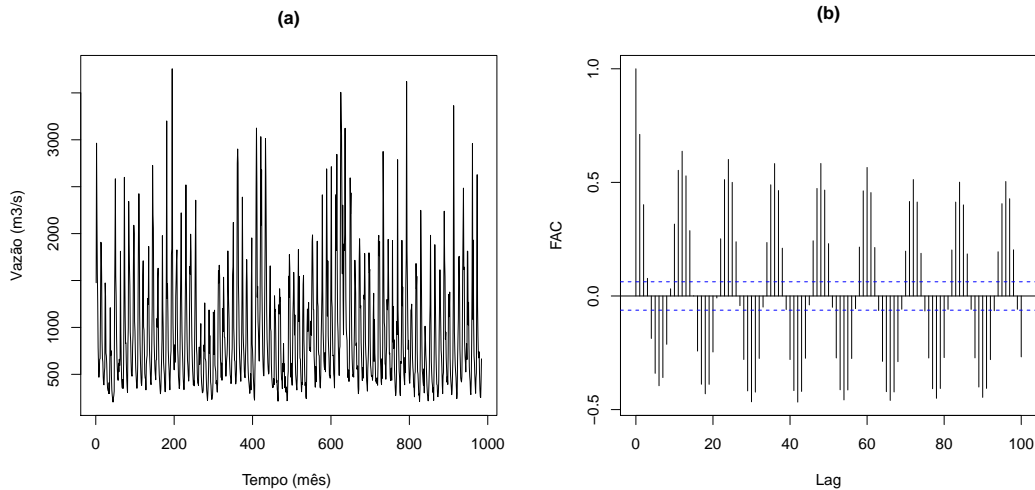


Figura 2.3: Série de vazão média mensal de Furnas em (a) e a função de autocorrelação em (b)

Por se tratar de uma série que apresenta comportamento sazonal, utilizamos a equação (2.4) para a obtenção do processo livre de sazonalidade, adotando  $r = 1, \dots, 82$ ,  $m = 1, \dots, 12$ ,  $\hat{\mu}_m$  e  $\hat{\sigma}_m$  são dados pelas respectivas equações (1.5) e (1.6). A utilização do valor  $S = 12$  é porque a série tem um ciclo hidrológico de 12 meses.

Para verificar se a série livre de sazonalidade apresenta tendência utilizamos o teste Cox-Stuart, proposto em Cox e Stuart (1955). Adotando o nível de significância de 5%, não existe evidências para rejeitarmos a hipótese de não haver tendência (p-valor de 0,8918). Utilizando o *software* R Core Team (2014), o teste pode ser calculado a partir



do pacote Caeiro e Mateus (2014)

As funções de autocorrelação e de autocorrelação parcial da série sem sazonalidade são apresentadas na Figura 2.4. Por meio dos gráficos notamos que a função de autocorrelação decai lentamente enquanto que a função de autocorrelação parcial apresenta valores diferentes de zero apenas nos três primeiros lags.

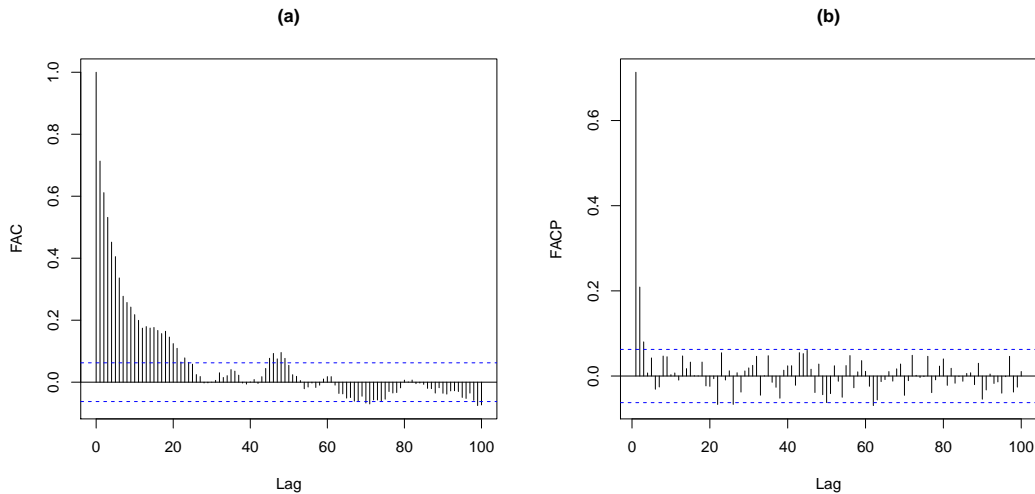


Figura 2.4: Funções de autocorrelação e de autocorrelação parcial da série livre de sazonalidade a partir da equação (2.4) em (a) e em (b), respectivamente

Para a escolha do melhor modelo para representar a série de vazões médias mensais ajustamos algumas ordens utilizando os modelos ARMA e NG-ARMA sem intercepto, os resultados estão apresentados na Tabela 2.5. Notamos pela mesma que o modelo que apresentou menor AIC e menor BIC em ambos os critérios foi o modelo NG-ARMA(1,1) sem intercepto.

As estimativas de máxima verossimilhança e os intervalos de confiança de 95% do modelo NG-ARMA(1,1) estão apresentados na Tabela 2.6. Por meio dos resultados observamos que todos os parâmetros são significativos e que a série apresenta caudas mais pesada do que as da distribuição normal, pois o valor de  $s$  é menor que 2, e notamos ainda que o intervalo de confiança não contém o valor 2, implicando assim que a distribuição normal não pode ser usada para explicar a série.

Através dos gráficos das funções de autocorrelação e autocorrelação parcial dos resíduos, que são apresentados na Figura 2.5, notamos que o ajuste do modelo NG-ARMA(1,1) eliminou a correlação dos resíduos. Também verificamos a aleatoriedade dos

Tabela 2.5: Seleção entre os modelos NG-ARMA e N-ARMA utilizando os critérios AIC e BIC para explicar a série de Furnas

Modelo	Log-Verossimilhança	AIC	BIC
NG-ARMA(2,0)	-961,96	1931,91	1951,43
NG-ARMA(3,0)	-958,24	1926,48	1950,88
NG-ARMA(0,1)	-1124,33	2254,67	2269,31
NG-ARMA(1,1)	<b>-956,67</b>	<b>1921,33</b>	<b>1940,85</b>
ARMA(3,0)	-999,76	2007,52	2027,04
ARMA(2,0)	-1001,81	2009,63	2024,27
ARMA(0,1)	-1174,87	2353,73	2363,49
ARMA(1,1)	-1000,52	2007,04	2021,67

Tabela 2.6: Estimativas pontual e intervalar dos parâmetros do modelo NG-ARMA(1,1) que melhor se ajustou a série de Furnas

Parâmetro	EMV	IC
$\phi_1$	0,877	(0,830; 0,924)
$\theta_1$	-0,358	(-0,456; -0,261)
$\sigma$	0,625	(0,543; 0,708)
$s$	1,212	(1,078; 1,346)

resíduos utilizando os testes Box-Pierce e Ljung-Box, mais detalhes sobre os testes podem ser encontrados em Brockwell e Davis (2002) e Harvey (1993), eles podem ser calculados a partir do *software* R Core Team (2014) utilizando Caeiro e Mateus (2014). Aplicando os testes nos resíduos, observamos que os resultados indicam que os resíduos são aleatórios.

Examinamos a adequabilidade da distribuição normal generalizada à série a partir do gráfico quantil teórico da distribuição normal generalizada versus quantil observado dos resíduos, o gráfico é apresentado na Figura 2.6. Observamos que os valores extremos dos resíduos divergem dos valores da distribuição normal generalizada, sendo que o mais grave ocorre para valores pequenos.

As previsões foram feitas para os meses do ano de 2012 considerando dois cenários: no 1º cenário adotamos que temos as observações até o mês de dezembro de 2011, para o 2º cenário assumimos que as previsões foram feitas apenas a um passo a frente, ou seja,

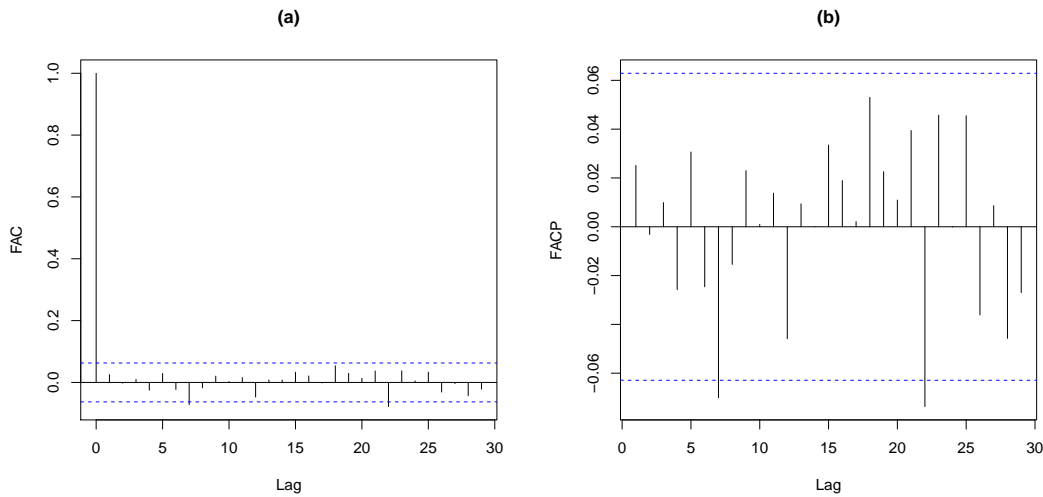


Figura 2.5: Funções de autocorrelação e autocorrelação parcial dos resíduos do ajuste do modelo NG-ARMA(1,1) à série de Furnas em (a) e em (b), respectivamente

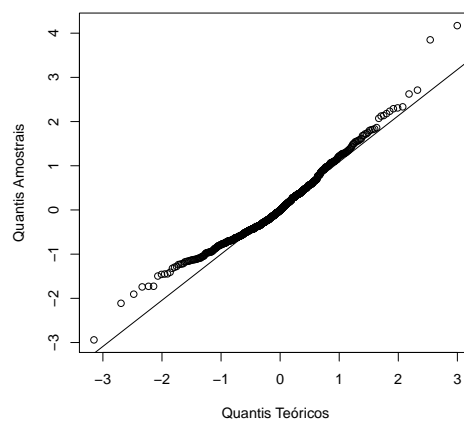


Figura 2.6: Gráfico quantil teórico da distribuição normal generalizada versus quantil observado dos resíduos do ajuste do modelo NG-ARMA(1,1) à série de Furnas

para prever o mês de janeiro de 2012 levamos em conta no ajuste do modelo as observações até dezembro de 2011, mas para prever o mês de fevereiro de 2012 reajustamos o modelo considerando as observações até o mês de janeiro de 2012, as previsões para os demais meses seguem essa mesma linha de raciocínio. Os valores preditos são apresentados na Figura 2.7.

Observando os dois cenários de predição apresentados na Figura 2.7 notamos que em ambos os cenários os erros de predições são maiores no início do ano. Esse resultado

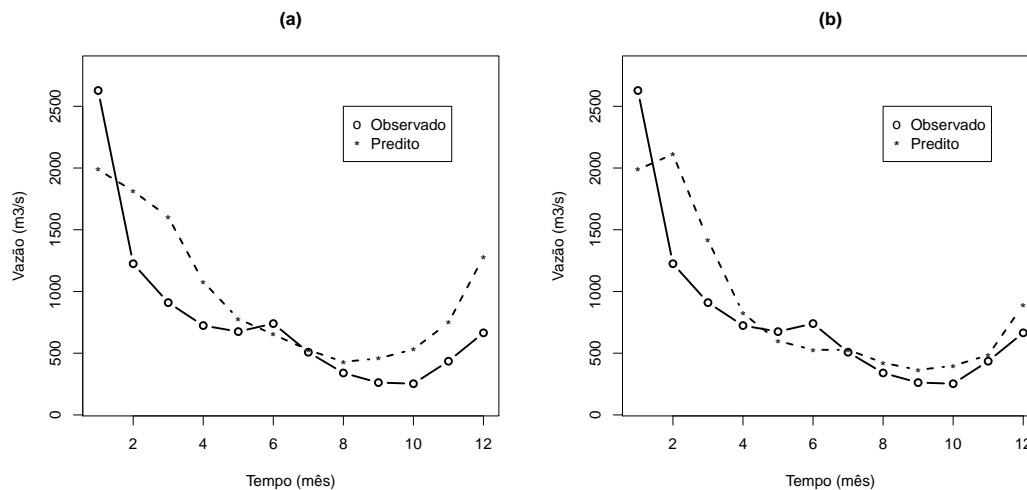


Figura 2.7: Previsões para a série de Furnas adotando o modelo NG-ARMA(1,1) e considerando o 1º cenário em (a), e o 2º cenário em (b)

já poderia ser esperado, pois são nos meses de janeiro e fevereiro que são apresentados os maiores volumes de vazão do ano, juntamente com a maior variabilidade.

Através da análise gráfica conseguimos notar que as previsões do segundo cenário se aproximam mais dos valores verdadeiros do que as previsões do primeiro cenário, também observamos esse resultado através dos critérios MAPE, EQM e EAM, que são apresentados na Tabela 2.7. Sabemos que essa comparação não é justa, pois no segundo cenário é utilizado mais informação do que no primeiro cenário, mas gostaríamos aqui de destacar que o resultado está em consonância com o estudo de simulação, porque no estudo de simulação vimos que previsões a um passo a frente produz erro de predição menor do que quando utilizamos previsões com vários passos a frente.

Tabela 2.7: Avaliação das previsões para a série de Furnas feitas a partir do modelo NG-ARMA(1,1) segundo os critérios MAPE, EQM e EAM

Cenário	MAPE(%)	EQM	EAM
1º	50,3	165.263,7	331,6
2º	31,4	134.174,0	254,2

### 2.3.2 Série da Média Mensal de Passageiros de Ônibus da Cidade de Iowa

Adotamos a série da média mensal de passageiros de ônibus da cidade de Iowa, localizada no estado de Iowa nos Estados Unidos. O estudo e previsões desta série é importante para o planejamento de futuras ações para o melhoria do transporte público coletivo.

A série foi observada de setembro de 1971 até dezembro de 1981, apresentada na Figura 2.8(a). Desconsideramos da análise da série as últimas 12 observações, ou seja, o ano de 1981, isso para que posteriormente possamos comparar as previsões com os verdadeiros valores. A série analisada possui 124 observações.

Podemos observar na Figura 2.8(a) que a série apresenta sazonalidade e tendência. Para a análise da série eliminamos primeiro a sazonalidade utilizando a equação (2.4), e depois retiramos a tendência, ainda existente, através da diferenciação da série, adotando a equação (2.19), assim, para o ajuste consideramos a configuração do modelo NG-ARMA para séries que apresentam sazonalidade com tendência, como apresentado na Seção 2.2. A série resultante é apresentada na Figura 2.8(b).

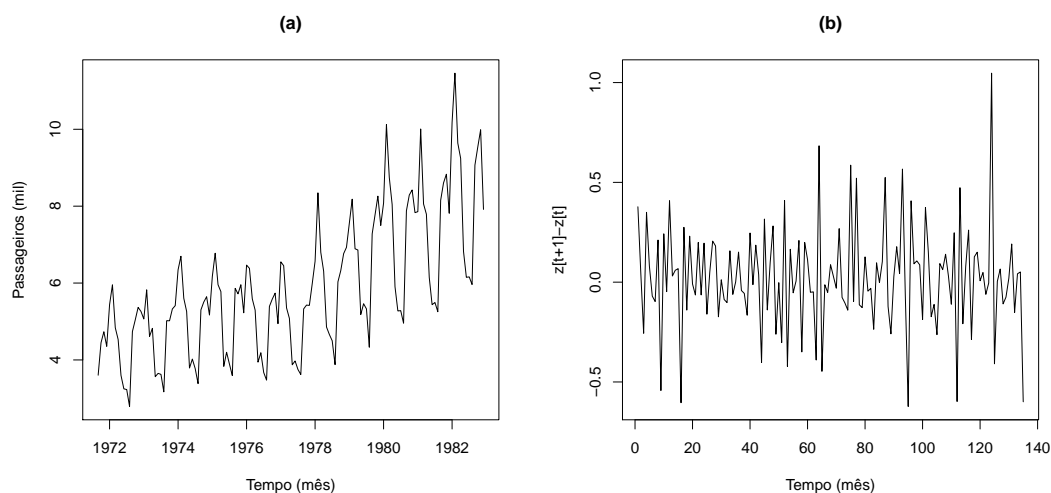


Figura 2.8: Série da média mensal de passageiros de ônibus da cidade de Iowa em (a) e série resultante livre de sazonalidade e tendência em (b)

Analisamos a correlação existente na série por intermédio das funções de autocorrelação e de autocorrelação parcial, que são apresentadas na Figura 2.9. Reconhecemos

através dos gráficos que possíveis modelos que melhor se ajustaram à série são os modelo NG-ARMA(1,0) ou NG-ARMA(0,1), pois as funções de autocorrelação e autocorrelação parcial apresentam valores diferentes de zero apenas no primeiro lag.

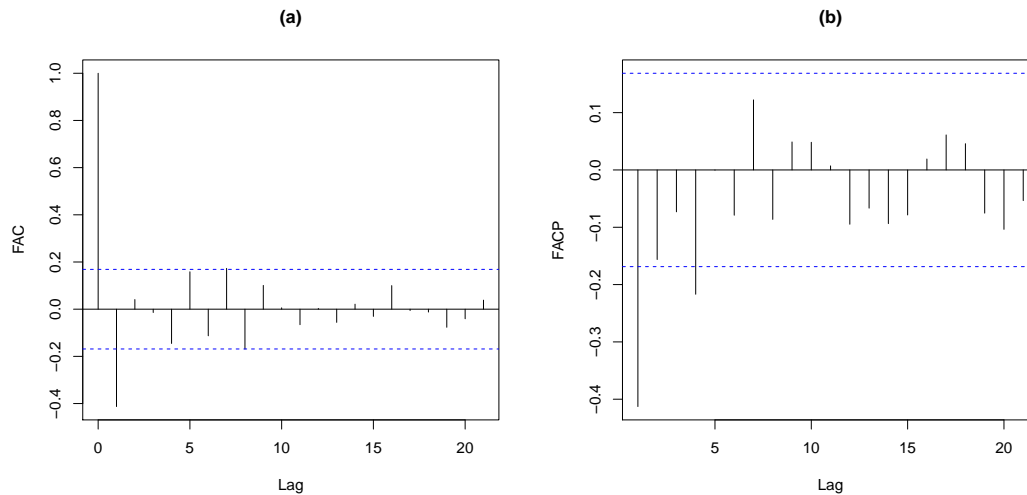


Figura 2.9: Funções de autocorrelação e de autocorrelação parcial da série livre de sazonalidade e de tendência em (a) e em (b), respectivamente

Adotamos os critérios AIC e BIC para as escolhas de qual distribuição utilizar, entre as distribuições normal e normal generalizada, e qual a melhor ordem. Ajustamos diferentes ordens com ambas distribuições, os resultados dos critérios são apresentados na Tabela 2.8.

Tabela 2.8: Seleção entre os modelos NG-ARMA e N-ARMA utilizando os critérios AIC e BIC para explicar a série de passageiros de ônibus da cidade de Iowa

Modelo	Log-Verossimilhança	AIC	BIC
NG-ARMA(2,0)	16,74	-25,49	-14,24
NG-ARMA(0,1)	18,09	<b>-30,19</b>	<b>-21,75</b>
NG-ARMA(0,2)	18,19	-28,38	-17,13
NG-ARMA(1,1)	18,09	-28,19	-16,94
ARMA(2,0)	14,37	-22,73	-14,29
ARMA(0,1)	15,58	-27,15	-21,53
ARMA(0,2)	15,67	-25,34	-16,90
ARMA(1,1)	15,58	-25,16	-16,72

Por meio dos resultados apresentados na Tabela 2.8 observamos que o modelo

que melhor explica a série em estudo é o NG-ARMA(0,1). Comparando a melhor ordem adotando a distribuição normal com a melhor ordem utilizando a distribuição normal generalizada, observamos que os valores dos critérios BIC ficaram mais próximos em comparação com os valores dos critérios AIC.

As estimativas de máxima verossimilhança e os intervalos de confiança dos parâmetros utilizando o modelo NG-ARMA(0,1) são apresentados na Tabela 2.9. Através dos resultados apresentados observamos que todos os parâmetros são significativos, e ainda notamos que o intervalo de confiança para o parâmetro  $s$  não contém o valor 2, indicando assim que a distribuição normal generalizada apresenta um ajuste melhor do que a distribuição normal.

Tabela 2.9: Estimativas pontual e intervalar dos parâmetros do modelo NG-ARMA(0,1) que melhor se ajusta à série de passageiros de ônibus da cidade de Iowa

Parâmetro	EMV	IC
$\theta_1$	-0,5162	(-0,6687;-0,3637)
$\sigma$	0,2162	(0,1296; 0,3027)
$s$	1,2993	(0,8061; 1,7926)

Analisamos a qualidade do ajuste do modelo NG-ARMA(0,1) à série de passageiros de ônibus por meio das funções de autocorrelação e autocorrelação parcial dos resíduos, que são apresentados na Figura 2.10. Notamos que os resíduos não são correlacionados. Aplicando os testes Box-Pierce e Ljung-Box observamos indícios que os resíduos são aleatórios, indicando assim que houve um bom ajuste do modelo à série.

De maneira análoga ao apresentado na Seção 2.3.1, verificamos a adequabilidade da distribuição normal generalizada à série a partir do gráfico quantil teórico da distribuição normal generalizada versus quantil observado dos resíduos, apresentado na Figura 2.11. Notamos que a maiores dos valores convergem para os valores da distribuição normal generalizada, mas nos extremos observamos alguns valores em divergência com os valores da distribuição normal generalizada.

Na Figura 2.12(a) apresentamos as previsões para a série transformada, calculamos aqui apenas previsões a um passo a frente, ou seja, adotamos a equação (2.22) e fazemos a atualização da série. Observe que para a previsão para a série original não

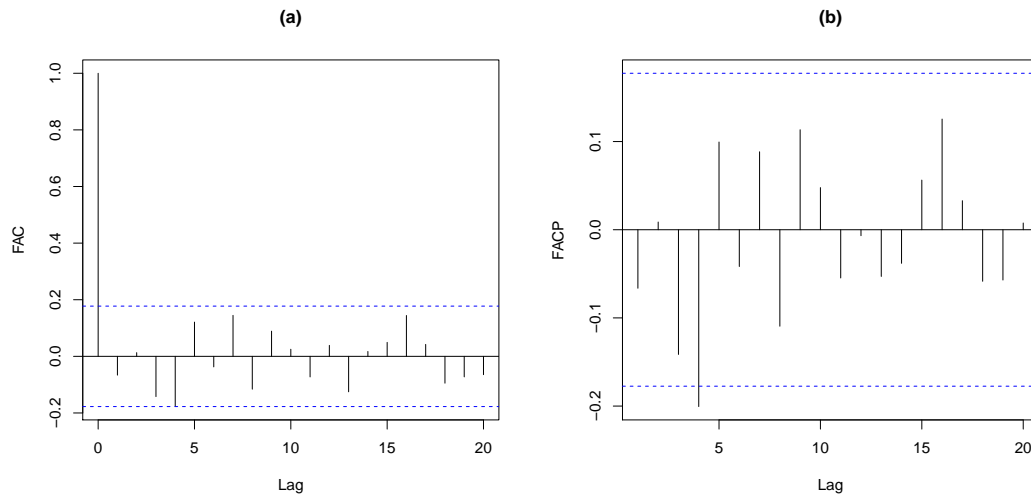


Figura 2.10: Funções de autocorrelação e autocorrelação parcial dos resíduos do ajuste do modelo NG-ARMA(0,1) à série de passageiros de ônibus da cidade de Iowa em (a) e em (b), respectivamente

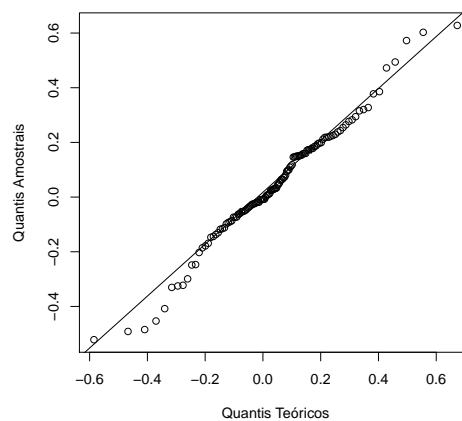


Figura 2.11: Gráfico quantil teórico da distribuição normal generalizada versus quantil observado dos resíduos do ajuste do modelo NG-ARMA(0,1) à série de passageiros de ônibus da cidade de Iowa

podemos apenas aplicar a equação (2.26), pois ajustamos a série livre de sazonalidade e



com uma diferenciação, portanto a equação de previsão para a série original é dada por

$$\begin{aligned}
 \hat{X}_t(1) = \hat{E}(X_{t+1}|\mathfrak{F}_t) &= \hat{E}(\hat{\mu}_{m_0} + \hat{\sigma}_{m_0} Z_{t+1}|\mathfrak{F}_t) \\
 &= \hat{\mu}_{m_0} + \hat{\sigma}_{m_0} \hat{E}(Z_{t+1}|\mathfrak{F}_t) \\
 &= \hat{\mu}_{m_0} + \hat{\sigma}_{m_0} \hat{E}(Y_t + Z_t|\mathfrak{F}_t) \\
 &= \hat{\mu}_{m_0} + \hat{\sigma}_{m_0} (\hat{Y}_t + Z_t),
 \end{aligned} \tag{2.32}$$

sendo que  $t + 1 = (r_0 + 1)S + m_0$  e  $\hat{Y}_t$  é a previsão para a série livre de sazonalidade e diferenciada. As previsões para a série original são apresentadas na Figura 2.12(b).

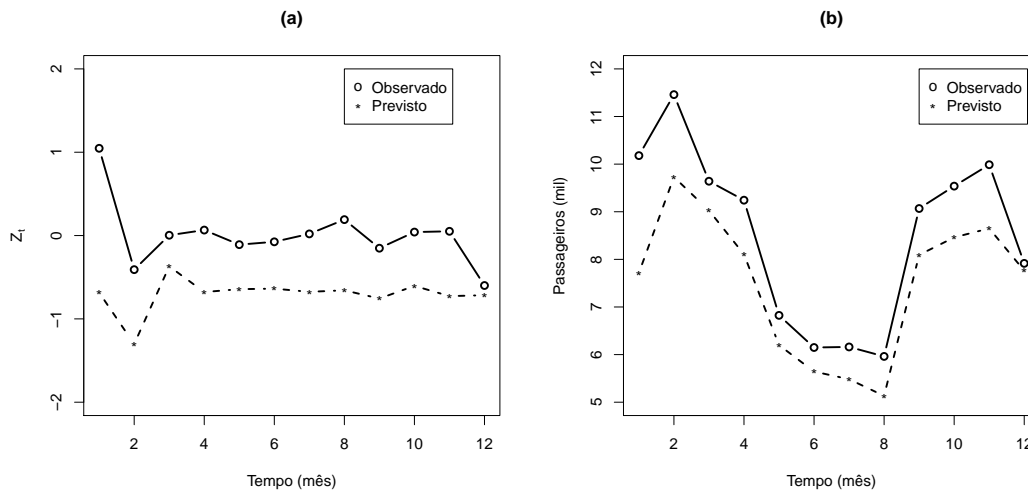


Figura 2.12: Previsão a um passo a frente para a série livre de sazonalidade e diferenciada em (a) e para a série de passageiros de ônibus da cidade de Iowa em (b)

Observamos na Figura 2.12(b) que as previsões em geral se apresentam próximas dos verdadeiros valores, mas nos meses de janeiro, fevereiro, outubro e novembro os valores preditos se mostraram mais longe dos verdadeiros valores em comparação ao restante do ano, notamos também que neste cenário as previsões subestimaram os valores observados. Para quantificar a qualidade das predições calculamos algumas métricas, que são apresentadas na Tabela 2.10.

Tabela 2.10: Avaliação das previsões para a série de passageiros de ônibus feitas a partir do modelo NG-ARMA(0,1) segundo os critérios MAPE, EQM e EAM

MAPE(%)	EQM	EAM
11,47	1,37	1,01

### 2.3.3 Série do Índice HICP dos Estados Unidos

Consideramos a série do índice HICP dos Estados Unidos observada mensalmente de dezembro de 1998 até maio de 2014, série apresentada na Figura 2.13(a). O índice HICP (do inglês Harmonised Index of Consumer Prices) é um importante índice que mede a inflação, para um melhor esclarecimento do índice HICP traduzimos o texto explicativo encontrado no site “eurostat.com” através do link “<http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/introduction>” e o apresentamos no parágrafo abaixo.

Os índices HICPs são indicadores econômicos construídos para medir as variações no tempo dos preços de bens de consumo e serviços adquiridos pelas famílias. Os HICPs dão medidas comparáveis de inflação na zona do euro, da União Europeia, do espaço econômico Europeu e de outros países, incluindo países aderentes e candidatos. Eles são calculados de acordo com uma abordagem harmonizada e um único conjunto de definições. Eles fornecem a medida oficial da inflação de preços ao consumidor na zona do euro, para efeitos de política monetária na área do euro e avaliar a convergência da inflação, conforme exigido nos termos dos critérios de Maastricht.

Como descrito no parágrafo anterior o índice HICP é um importante parâmetro econômico, portanto seu estudo e previsões de seus valores são de grande interesse de governos, de economistas e da sociedade como um todo.

Neste estudo desconsideramos as últimas doze observações, ou seja, as observações de junho de 2013 a maio de 2014, para que futuramente possamos comparar as previsões obtidas a partir do modelo ajustado com as verdadeiras observações. Diferenciamos a série temporal para torna-lá estacionária, a série diferenciada é apresentada na Figura 2.13(b).

Investigamos a correlação existente na série diferenciada através das funções de autocorrelação e autocorrelação parcial, estas são apresentadas na Figura 2.14. Observamos a existência de correlação nos primeiros lags e também em torno dos lags 12 e 13, indicando assim uma correlação sazonal, com isso notemos a necessidade de um componente sazonal na modelagem.

O modelo NG-ARMA com componente sazonal, definido na Seção 2.2, será ajustado à série temporal diferenciada. Utilizamos os critérios AIC e BIC para as escolhas

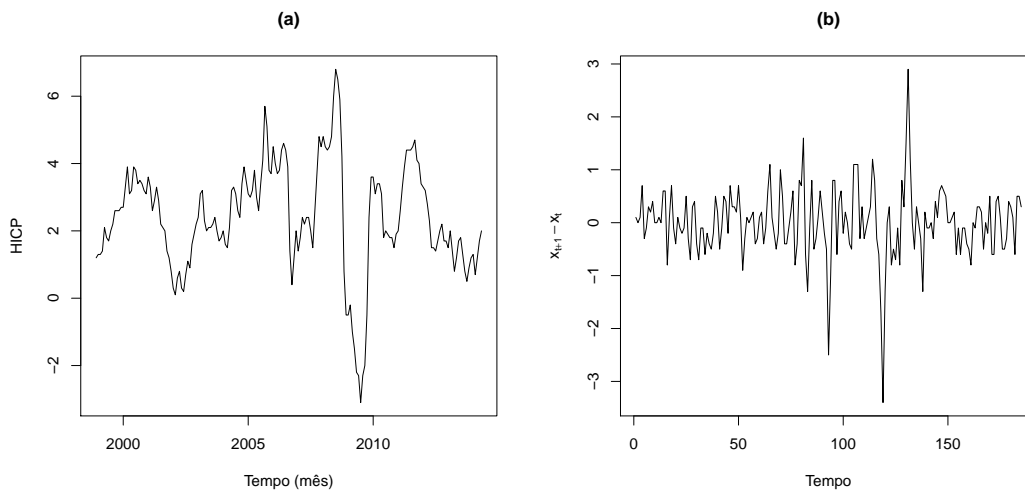


Figura 2.13: Série do índice HICP em (a) e série do índice HICP diferenciada em (b)

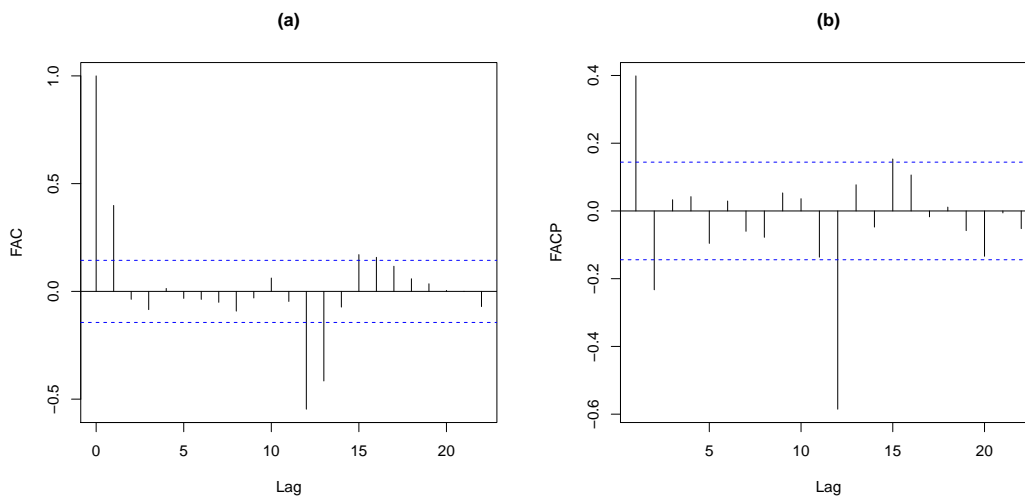


Figura 2.14: Funções de autocorrelação e autocorrelação parcial da série diferenciada do índice HICP em (a) e (b), respectivamente

da melhor distribuição, entre as distribuições normal e normal generalizada, e da melhor ordem. A escolha do melhor modelo é decidida entre vários ajustes, para isso adotamos diferentes ordens, sendo considerado também modelos com parâmetros autorregressivos (AR) e de médias móveis (MA) com defasagem de um ano ou mais. Os resultados dos critérios são apresentados na Tabela 2.11.

A partir dos resultados mostrados na Tabela 2.11 observamos que o melhor modelo segundo ambos os critérios é o NG-ARMA(2,0)S(1,1), ou seja, a distribuição normal generalizada se sobressai em relação à normal e o modelo apresenta componente sazonal

Tabela 2.11: Seleção entre os modelos NG-ARMA e N-ARMA utilizando os critérios AIC e BIC para explicar à série do índice HICP

Modelo	Log-Verossimilhança	AIC	BIC
NG-ARMA(2,0)S(0,1)	-89,13	188,27	204,03
NG-ARMA(2,0)S(1,1)	-85,02	<b>182,04</b>	<b>200,96</b>
NG-ARMA(2,0)S(1,0)	-106,63	223,27	239,03
N-ARMA(2,0)S(0,1)	-96,13	200,26	212,87
N-ARMA(2,0)S(1,1)	-91,88	193,77	209,54
N-ARMA(2,0)S(1,0)	-115,25	238,50	251,12

com defasagem de um ano de ordem (1,1). As estimativas de máxima verossimilhança e os intervalos de confiança dos parâmetros do modelo escolhido por meio dos critérios são exibidos na Tabela 2.12.

Tabela 2.12: Estimativas pontual e intervalar dos parâmetros do modelo NG-ARMA(2,0)S(1,1) que melhor se ajustou à série do índice HICP

Parâmetro	EMV	IC
$\phi_1$	0,4430	(0,2944; 0,5915)
$\phi_2$	-0,2240	(-0,3488; -0,0993)
$\phi_1^*$	-0,2370	(-0,2980; -0,1761)
$\theta_1^*$	-0,7289	(-0,8250; -0,6327)
$\sigma$	0,3631	(0,2196; 0,5067)
$s$	1,1233	(0,7795; 1,4671)

Através dos resultados expostos na Tabela 2.12 observamos que todos os parâmetros são significativos no nível de significância de 5%, o intervalo de confiança do parâmetro  $s$  não contém o valor dois.

Verificamos a adequabilidade do ajuste por meio das funções de autocorrelação e autocorrelação parcial, apresentadas na Figura 2.15, e dos testes Box-Pierce e Ljung-Box. Notamos a não existência de correlação nos resíduos, indicando assim que houve um bom ajuste do modelo à série.

Investigamos a adequabilidade da distribuição normal generalizada à série a partir do gráfico quantil teórico da distribuição normal generalizada versus quantil observado

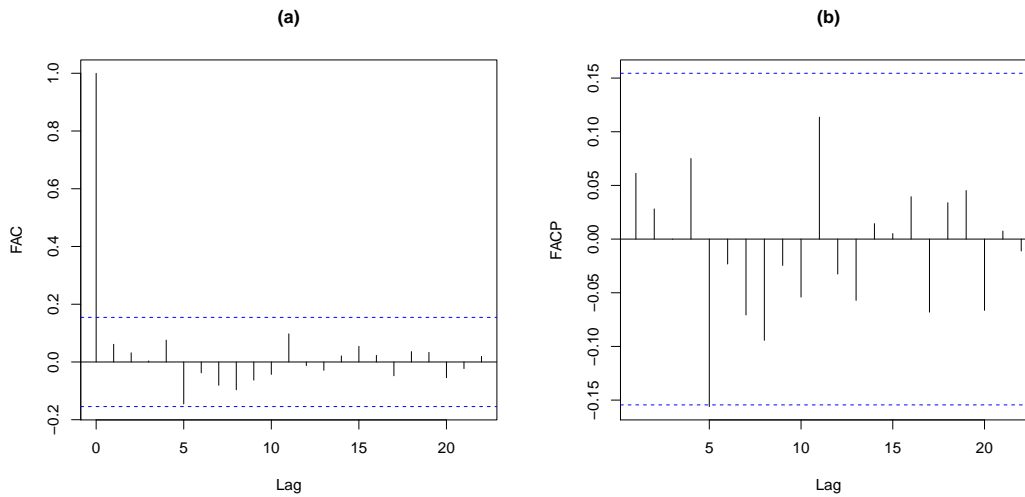


Figura 2.15: Funções de autocorrelação e autocorrelação parcial dos resíduos do ajuste do modelo NG-ARMA(2,0)S(1,1) à série do índice HICP dos Estados Unidos em (a) e em (b), respectivamente

dos resíduos, mostrado na Figura 2.16. Notamos que a maioria dos valores dos resíduos apresentam comportamento similar ao comportamento da distribuição normal generalizada.

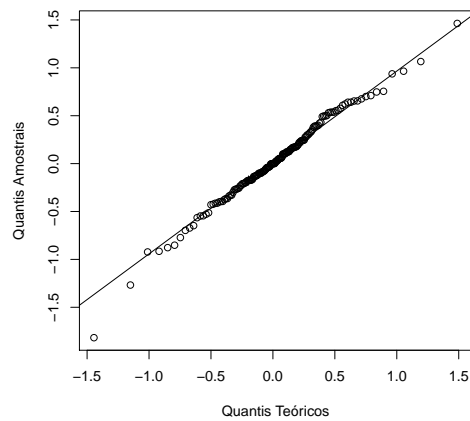


Figura 2.16: Gráfico quantil teórico da distribuição normal generalizada versus quantil observado dos resíduos do ajuste do modelo NG-ARMA(2,0)S(1,1) à série do índice HICP

Para a predição da série original utilizamos a equação dada por

$$\begin{aligned} \hat{X}_t(1) &= \hat{E}(X_{t+1} | \mathfrak{F}_t) = \hat{E}(Y_t + X_t | \mathfrak{F}_t) \\ &= \hat{Y}_t + X_t, \end{aligned} \quad (2.33)$$

sendo que  $\hat{Y}_t$  é a predição para a série diferenciada. Portanto a partir da equação (2.33) e das predições da série diferenciada obtemos as predições para a série original, para os dois cenários adotados na Seção 2.3.1, que são apresentadas na Figura 2.17.

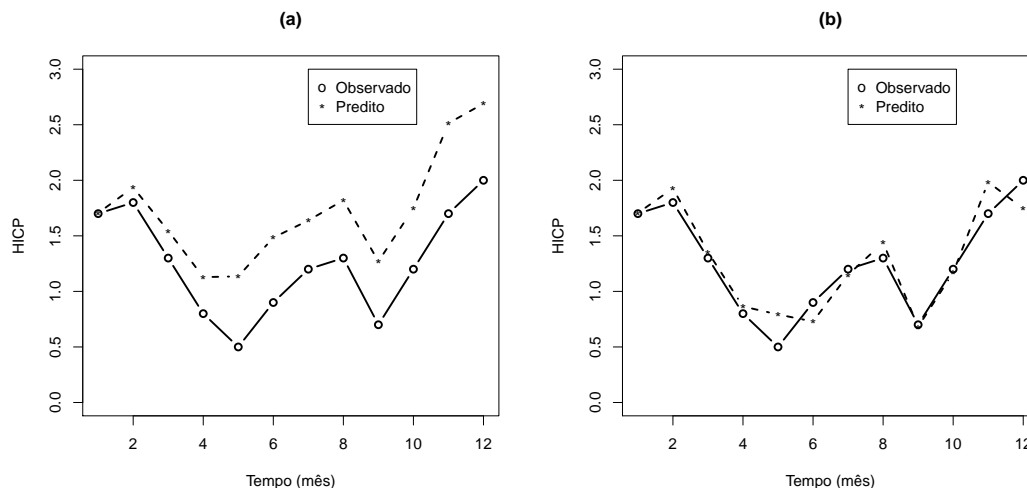


Figura 2.17: Previsões para a série HICP adotando o modelo NG-ARMA(2,0)S(1,1) considerando o 1º cenário em (a), e o 2º cenário em (b)

Verificamos por meio da Figura 2.17 que as previsões para o primeiro cenário vão se distanciando dos verdadeiros valores, este comportamento já era esperado, pois não utilizamos novas informações. No segundo cenário as previsões se apresentaram próximas dos verdadeiros valores, sendo que nos meses de maio, novembro e dezembro foram observados os valores mais discrepantes.

Quantificamos a qualidade das previsões por meio de três métricas, apresentadas na Tabela 2.13. Observamos que as métricas apresentam um baixo valor para o segundo cenário.

Tabela 2.13: Avaliação das previsões para a série do índice HICP feitas a partir do modelo NG-ARMA(2,0)S(1,1) segundo os critérios MAPE, EQM e EAM

Cenário	MAPE(%)	EQM	EAM
1º	45,47	0,26	0,46
2º	12,22	0,03	0,12

## 2.4 Conclusão

Neste capítulo propusemos um novo modelo para séries temporais que apresentam distribuição dos dados com caudas mais leves ou mais pesadas do que as da distribuição normal. Algumas importantes propriedades do novo modelo foram apresentadas, juntamente com um método de estimação utilizando a função de máxima verossimilhança e também um método de previsão. Estudos de simulação para verificar as propriedades frequentistas e escolha da ordem pelos critérios AIC e BIC apresentaram ótimos resultados.

Através das aplicações com as séries de vazões médias mensais da usina hidrelétrica de Furnas, da quantidade de passageiros de ônibus da cidade de Iowa e do índice HICP dos Estados Unidos, mostramos que o modelo NG-ARMA é uma alternativa ao modelo ARMA com distribuição normal quando os dados apresentam curtose diferente da distribuição normal.

# Capítulo 3

## Modelo Assimétrico Normal Generalizada ARMA

O modelo NG-ARMA, apresentado no Capítulo 2, é uma alternativa para modelar séries temporais que apresentam simetria e curtose diferente da distribuição normal. Com o interesse em estender este modelo para que possa ser uma alternativa para modelar séries temporais onde a suposição de simetria não seja verificada, propomos duas alternativas, a primeira é a utilização da distribuição log-normal generalizada e a segunda é a adoção da transformação Box-Cox.

Neste capítulo apresentamos os modelos log-normal generalizada ARMA (LNG-ARMA) e o Box-Cox normal generalizada ARMA (BC-NG-ARMA), que podem ser utilizados para explicar séries temporais que apresentam assimetria. Na Seção 3.1 apresentamos o modelo LNG-ARMA, descrevemos uma maneira para fazer previsões e aplicamos o modelo proposto na série de vazões médias mensais afluentes da usina hidrelétrica de Furnas. O modelo BC-NG-ARMA é apresentado na Seção 3.2 juntamente com um método de previsão, a construção da função de verossimilhança e a aplicação do modelo proposto para explicar as séries de vazões das usinas de Furnas e de Sobradinho. As conclusões são apresentadas na Seção 3.3.



### 3.1 Modelo Log-Normal Generalizada ARMA

Uma variável aleatória  $Y$  tem distribuição log-normal generalizada se a transformação  $X = \log(Y)$  tem distribuição normal generalizada. A função densidade de probabilidade da variável aleatória  $Y$  é dada por

$$f(y|\mu, \sigma, s) = \frac{s}{2y\sigma\Gamma(1/s)} \exp\left(-\left|\frac{\log(y) - \mu}{\sigma}\right|^s\right), \quad (3.1)$$

sendo que  $y > 0$ ,  $-\infty < \mu < \infty$ ,  $\sigma > 0$  e  $s > 0$ . Algumas formas da função densidade de probabilidade da distribuição log-normal generalizada são apresentadas na Figura 3.1. Alguns trabalhos que utilizam a distribuição log-normal generalizada são Singh *et al.* (2012), Gupta e Lvin (2005) e Martín e Pérez (2009), sendo os dois primeiros na área de análise de sobrevivência.

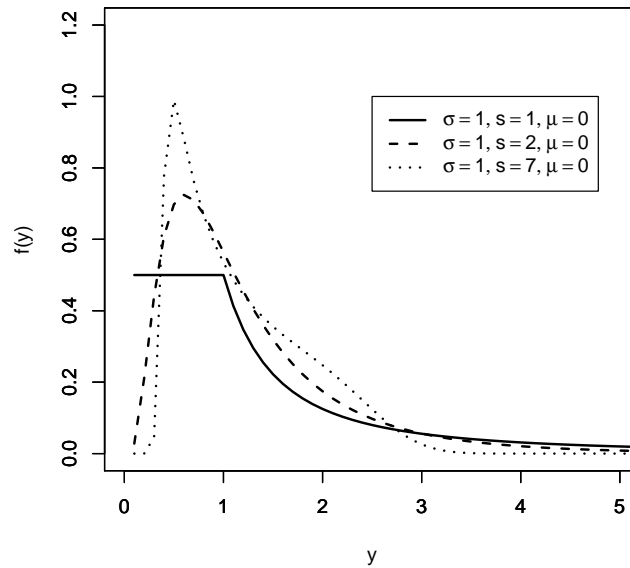


Figura 3.1: Algumas formas da função densidade da distribuição log-normal generalizada

Uma extensão natural do modelo NG-ARMA para considerar um caso específico de assimetria é considerar a distribuição log-normal generalizada em vez da distribuição normal generalizada, obtendo-se então o modelo log-normal generalizada ARMA (LNG-ARMA). Considere  $\{X_{t(r,m)}; r \geq 1; m = 1, \dots, S\}$  um processo estocástico sazonal com distribuição condicional log-normal generalizada, ou seja,

$$f_{X_t|\mathfrak{F}_{t-1}}(x_t|\mu, \sigma, s) = \frac{s}{2x_t\sigma\Gamma(1/s)} \exp\left(-\left|\frac{\log(x_t) - \mu_t}{\sigma}\right|^s\right). \quad (3.2)$$

O processo  $\{Z_t, t \geq 1\}$  obtido a partir da equação

$$Z_t = \frac{\log(X_{t(r,m)}) - \mu_m}{\sigma_m}, \quad (3.3)$$

sendo que  $\mu_m = E\{\log(X_{t(r,m)})\}$  e  $\sigma_m = \sqrt{\text{Var}\{\log(X_{t(r,m)})\}}$  é livre de sazonalidade. Notemos ainda que a distribuição condicional do processo  $\{Z_t, t \geq 1\}$  depende do logaritmo da distribuição condicional do processo  $\{X_{t(r,m)}; r \geq 1; m = 1, \dots, S\}$ , que tem distribuição log-normal generalizada, logo a distribuição condicional do processo  $\{Z_t, t \geq 1\}$  tem distribuição normal generalizada, portanto tudo o que foi feito no Capítulo 2 vale para o processo  $\{Z_t, t \geq 1\}$ . Ou seja, para adotarmos a distribuição log-normal generalizada como sendo a distribuição condicional da série temporal, basta aplicarmos o logaritmo na série temporal e ajusta-lá com a distribuição normal generalizada.

### 3.1.1 Previsão

Estamos interessados em prever um valor de  $X_{t+h}$  com  $h \geq 1$ , supondo que a série foi observada até o instante  $t$ , para isso precisamos inicialmente calcular as previsões para a série  $Z_t$ . A partir das observações da série original conseguimos obter a série  $\{z_1, z_2, \dots, z_t\}$ , até o instante  $t$ , o valor  $\hat{Z}_t(h) = \hat{E}(Z_{t+h}|\mathfrak{F}_t)$  será a previsão de *origem*  $t$  e *horizonte*  $h$ . Como  $Z_t$  tem distribuição condicional normal generalizada então para o modelo de ordem  $p$  e  $q$  a previsão para  $h = 1$  é feita através da equação dada em (2.22) e para  $h > 1$  temos que a previsão é dada por (2.25).

Para que não precisemos fazer aproximações optamos por prever a série original através da mediana, para isso definimos

$$P(Z_{t+h} \leq z_{t+h}^{(M)}|\mathfrak{F}_t) = 0,5,$$

assim  $z_{t+h}^{(M)}$  é o valor da mediana da distribuição  $Z_{t+h}|\mathfrak{F}_t$ , mas podemos reescrever a série  $Z_t$  como função da série  $X_t$ , disto temos que

$$P\left(\frac{\log(X_{t(r_1, m_1)}) - \hat{\mu}_{m_1}}{\hat{\sigma}_{m_1}} \leq z_{t+h}^{(M)}|\mathfrak{F}_t\right) = 0,5,$$

em que  $t(r, m) + h = t(r_1, m_1)$ . Definindo  $\mu_t^*(h)$  como a previsão para  $X_{t(r,m)+h}$  temos que

$$\mu_t^*(h) = \exp[z_{t+h}^{(M)}\hat{\sigma}_{m_1} + \hat{\mu}_{m_1}],$$

como a distribuição normal generalizada é simétrica em torno da média, assim a média é igual a mediana, disto aproximamos a mediana pela previsão da série sem sazonalidade, resultando em

$$\mu_t^*(h) = \exp[\widehat{\mu}_{t+h}\widehat{\sigma}_{m_1} + \widehat{\mu}_{m_1}], \quad (3.4)$$

para qualquer  $h \geq 1$ .

Portanto, mesmo ajustando o modelo NG-ARMA à série transformada conseguimos fazer previsões para a série original. Mas observamos que antes devemos calcular as previsões para a série transformada.

### 3.1.2 Aplicação à Série de Furnas

Para exemplificar a teoria desenvolvida nesta Seção, utilizamos a série de vazões médias mensais afluentes ( $m^3/s$ ) da usina hidrelétrica de Furnas apresentada na Seção 2.3.1. Por se tratar de uma série que apresenta apenas valores positivos e sazonalidade, como pode ser visto na Figura 2.3, para o ajuste utilizamos a série obtida a partir da equação (3.3), sendo que  $r = 1, \dots, 82$ ,  $m = 1, \dots, 12$ .

Ajustamos algumas ordens utilizando os modelos LN-ARMA e LNG-ARMA sem intercepto, e a escolha do melhor modelo para representar a série de vazões médias mensais foi feita utilizando os critérios AIC e BIC, os resultados estão apresentados na Tabela 3.1. Notamos que o modelo que apresentou menor AIC e menor BIC foi o modelo LNG-ARMA(1,1) sem intercepto, em ambos os critérios.

Os resultados do ajuste do modelo LNG-ARMA(1,1) estão apresentados na Tabela 3.2, lembrando que os intervalos de confiança foram construídos assumindo normalidade assintótica e nível de confiança de 95%. Observando os intervalos de confiança notamos que todos os parâmetros são significativos e a partir das estimativas de máxima verossimilhança constatamos que a distribuição da série apresenta caudas mais pesada do que as da distribuição normal, pois o valor de  $s$  é menor que 2.

Através dos gráficos das funções de autocorrelação e autocorrelação parcial, apresentados na Figura (3.2), e dos testes Box-Pierce e Ljung-Box investigamos a correlação dos resíduos. Observamos indícios que os resíduos não são correlacionados tanto pelas funções quanto pelos testes.

Tabela 3.1: Seleção de modelo utilizando os critérios AIC e BIC

Modelo	Log-Verossimilhança	AIC	BIC
LNG-ARMA(2,0)	-999,09	2006,19	2025,70
LNG-ARMA(3,0)	-997,19	2004,38	2028,77
LNG-ARMA(0,1)	-1168,02	2342,04	2356,68
LNG-ARMA(1,1)	<b>-996,35</b>	<b>2000,68</b>	<b>2020,22</b>
LN-ARMA(3,0)	-1004,39	2016,78	2036,30
LN-ARMA(2,0)	-1007,80	2021,60	2036,24
LN-ARMA(0,1)	-1173,05	2350,10	2359,86
LN-ARMA(1,1)	-1007,80	2021,60	2036,24

Tabela 3.2: Estimativas do modelo LNG-ARMA(1,1)

Parâmetro	EMV	IC
$\phi_1$	0,869	(0,825; 0,912)
$\theta_1$	-0,307	(-0,389; -0,224)
$\sigma$	0,776	(0,686; 0,865)
$s$	1,458	(1,275; 1,652)

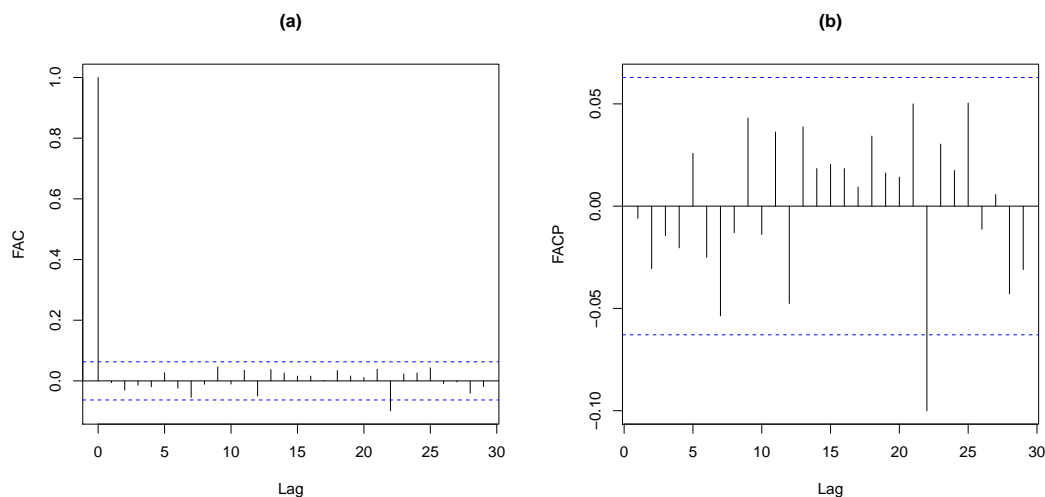


Figura 3.2: Funções de autocorrelação e autocorrelação parcial dos resíduos do ajuste do modelo LNG-ARMA(1,1) à serie de Furnas em (a) e em (b), respectivamente

Testamos a adequabilidade da distribuição normal generalizada analisando o gráfico quantil teórico da distribuição normal generalizada versus quantil observado dos resí-

duos, apresentado na Figura 3.3. Por ele notamos que a distribuição dos resíduos apresenta um comportamento muito parecido com a distribuição normal generalizada.

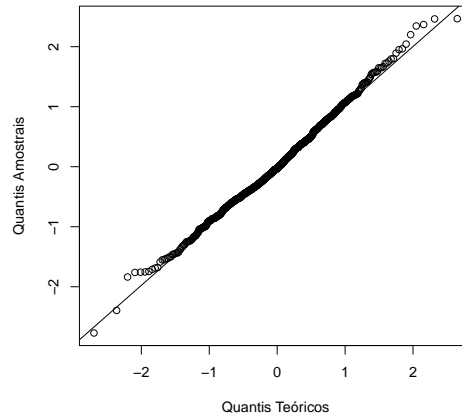


Figura 3.3: Gráfico quantil teórico da distribuição normal generalizada versus quantil observado

A partir das previsões para a série transformada e utilizando a equação (3.4) apresentamos as previsões para a série original para o ano de 2012 na Figura 3.4, em (a) as previsões para o 1º cenário e em (b) as previsões para o 2º cenário, considerando os dois cenários definidos na Seção 2.3.1

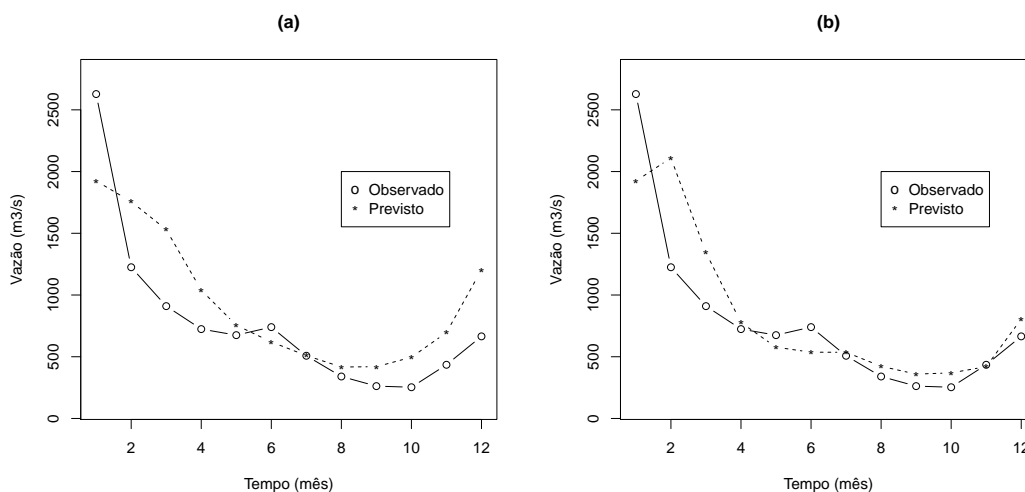


Figura 3.4: Previsões para a série de Furnas considerando o 1º cenário em (a) e o 2º cenário em (b)

Notamos através da Figura 3.4 que para os três primeiros meses do ano de 2012 as previsões ficaram mais distantes dos valores observados em comparação aos outros meses,

isso em ambos os cenários. Já no 2º cenário, as previsões de um modo geral se apresentam próximas dos valores observados principalmente a partir do mês de abril.

Através da Tabela 3.3 observamos os resultados dos critérios MAPE, EQM e EAM para as previsões. Comparando estes resultados com os obtidos através das previsões na Seção 2.3.1, notamos que os resultados obtidos através do ajuste com a distribuição log-normal generalizada apresentam menor erro de previsão do que quando utilizamos a distribuição normal generalizada, isso para os três critérios adotados.

Tabela 3.3: Resumos dos critérios MAPE, EQM e EAM para a série de Furnas

Cenário	MAPE(%)	EQM	EAM
1º	44,2	144.289,2	304,2
2º	27,8	130.952,8	238,1

## 3.2 Modelo Box-Cox Normal Generalizada ARMA

Box e Cox (1964) propuseram uma transformação, conhecida como transformação de Box-Cox, que tem como objetivo a transformação dos dados para que o modelo linear e homocedástico seja adequado, ou seja, tornar os dados com distribuição aproximadamente normal e com variância constante. Com basicamente os mesmo objetivos, mas agora no contexto de séries temporais, aplicamos a transformação Box-Cox para tentarmos obter uma série resultante mais simétrica do que a original, que tenha distribuição aproximadamente normal generalizada e que seja homocedástica. Seja então  $\{X_{t(r,m)}; r \geq 1; m = 1, \dots, S\}$  uma série temporal sazonal observada. Aplicando a transformação Box-Cox na série temos

$$Y_{t(r,m)}^{(\lambda)} = \begin{cases} \frac{X_{t(r,m)}^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(X_{t(r,m)}), & \lambda = 0 \end{cases} \quad (3.5)$$

sendo  $\{Y_{t(r,m)}^{(\lambda)}; r \geq 1; m = 1, \dots, S\}$  a série resultante. Eliminamos a sazonalidade pertencente a série utilizando a equação dada por

$$Z_t^{(\lambda)} = \frac{Y_{t(r,m)}^{(\lambda)} - \mu_m}{\sigma_m}, \quad (3.6)$$

sendo que  $\mu_m = E\{Y_{t(r,m)}^{(\lambda)}\}$ ,  $\sigma_m = \sqrt{Var\{Y_{t(r,m)}^{(\lambda)}\}}$ ,  $r \geq 1$  e  $m = 1, 2, \dots, S$ . Assumindo que a distribuição condicional da série  $\{Z_t^{(\lambda)}; t \geq 1\}$  dada toda a informação passada do processo possui função densidade de probabilidade normal generalizada, ou seja, adotamos que  $Z_t^{(\lambda)} | \mathfrak{F}_{t-1} \sim \text{NG}(\mu_t, \sigma, s)$  sendo que

$$\mu_t = \sum_{j=1}^p \phi_j z_{t-j}^{(\lambda)} + \sum_{j=1}^q \theta_j (z_{t-j}^{(\lambda)} - \mu_{t-j}), \quad (3.7)$$

portanto usamos a transformação Box-Cox para que a série transformada tenha distribuição aproximadamente normal generalizada e variância constante. O modelo definido pelas equações (3.5), (3.6) e (3.7) é chamado de Box-Cox Normal Generalizada ARMA (BC-NG-ARMA).

### 3.2.1 Estimação de Máxima Verossimilhança

A estimação de máxima verossimilhança será feita em dois passos. No primeiro passo iremos estimar o parâmetro  $\lambda$ , que definimos como parâmetro perturbador, e no segundo passo estimamos o vetor  $(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s)$  sendo que  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$  e  $\sigma$  e  $s$  são escalares, sendo estes os parâmetros de interesse.

Sejam  $\{x_1, \dots, x_n\}$  a série temporal observada e  $\{z_t^{(\lambda)}; t \geq 1\}$  a série transformada e livre de sazonalidade, para algum valor de  $\lambda$  em (3.5). A função de verossimilhança para o vetor de parâmetros de interesse condicional ao valor de  $\lambda$  é dada por

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \lambda, \mathbf{x}) = \prod_{j=r^*+1}^n f_{Z_t^{(\lambda)}}(z) J_{ZX},$$

sendo que  $r^* = \max(p, q)$  e  $J_{ZX} = \frac{\partial Z}{\partial X}$ . Com  $\lambda \neq 0$  observe que

$$z_t^{(\lambda)} = \frac{x_{t(r,m)}^{\lambda-1} - \mu_m}{\sigma_m}$$

então

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \lambda, \mathbf{x}) = \left[ \prod_{r=1}^N \prod_{m=1}^S \frac{x_{t(r,m)}^{\lambda-1}}{\sigma_m} \right] \left( \frac{s}{2\sigma\Gamma(1/s)} \right)^{n-r^*} \times \exp \left\{ -\frac{1}{\sigma^s} \sum_{t=r^*+1}^n \left| z_t^{(\lambda)} - \sum_{j=1}^p \phi_j z_{t-j}^{(\lambda)} - \sum_{j=1}^q \theta_j (z_{t-j}^{(\lambda)} - \mu_{t-j}) \right|^s \right\}, \quad (3.8)$$

mas caso  $\lambda = 0$  então

$$z_t^{(\lambda)} = \frac{\log(x_{t(r,m)}) - \mu_m}{\sigma_m},$$

logo

$$\begin{aligned} L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \lambda, \mathbf{x}) &= \left[ \prod_{r=1}^N \prod_{m=1}^S \frac{1}{x_{t(r,m)} \sigma_m} \right] \left( \frac{s}{2\sigma\Gamma(1/s)} \right)^{n-r^*} \times \\ &\exp \left\{ -\frac{1}{\sigma^s} \sum_{t=r^*+1}^n \left| z_t^{(\lambda)} - \sum_{j=1}^p \phi_j z_{t-j}^{(\lambda)} - \sum_{j=1}^q \theta_j (z_{t-j}^{(\lambda)} - \mu_{t-j}) \right|^s \right\}, \end{aligned} \quad (3.9)$$

note que nas equações (3.8) e (3.9) temos  $\sigma$  e  $\sigma_m$ , sendo que  $\sigma_m$  é estimado como na equação (3.6) e  $\sigma$  é estimado na função de verossimilhança.

Seja  $l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \lambda, \mathbf{x}) = \log[L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \lambda, \mathbf{x})]$ , então a função log-verossimilhança condicional quando  $\lambda \neq 0$  é dada por

$$\begin{aligned} l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \lambda, \mathbf{x}) &= (\lambda - 1) \sum_{r=1}^N \sum_{m=1}^S \log(x_{t(r,m)}) - N \sum_{m=1}^S \log(\sigma_m) + (n - r^*) \log \left( \frac{s}{2\sigma\Gamma(1/s)} \right) \\ &\quad - \frac{1}{\sigma^s} \sum_{t=r^*+1}^n \left| z_t^{(\lambda)} - \sum_{j=1}^p \phi_j z_{t-j}^{(\lambda)} - \sum_{j=1}^q \theta_j (z_{t-j}^{(\lambda)} - \mu_{t-j}) \right|^s, \end{aligned}$$

e quando  $\lambda = 0$  a função log-verossimilhança condicional é dada por

$$\begin{aligned} l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \lambda, \mathbf{x}) &= - \sum_{r=1}^N \sum_{m=1}^S \log(x_{t(r,m)}) - N \sum_{m=1}^S \log(\sigma_m) + (n - r^*) \log \left( \frac{s}{2\sigma\Gamma(1/s)} \right) \\ &\quad - \frac{1}{\sigma^s} \sum_{t=r^*+1}^n \left| z_t^{(\lambda)} - \sum_{j=1}^p \phi_j z_{t-j}^{(\lambda)} - \sum_{j=1}^q \theta_j (z_{t-j}^{(\lambda)} - \mu_{t-j}) \right|^s. \end{aligned}$$

Para o cálculo das estimativas de máxima verossimilhança adotamos a utilização de métodos iterativos devido à complexidade da função de máxima verossimilhança. Neste trabalho adotamos o método “L-BFGS-B” que está disponível no *software* R Core Team (2014) (Byrd *et al.*, 1995). Os intervalos de confiança assintóticos para os parâmetros de interesse são obtidos considerando os estimadores de máxima verossimilhança, a inversa da matriz de informação observada de Fisher e que os estimadores apresentam normalidade assintótica.

Observe que a função de máxima verossimilhança descrita anteriormente é condicional ao parâmetro  $\lambda$ , com isso antes de estimarmos os parâmetros de interesse temos



que inferir sobre o parâmetro perturbador. A estimativa de máxima verossimilhança do parâmetro  $\lambda$  é obtida a partir da função de verossimilhança perfilada, ou seja, previamente determinamos um conjunto de possíveis valores para  $\lambda$ , para cada valor deste conjunto encontramos o valor do máximo da função log-verossimilhança condicional. O valor escolhido para  $\lambda$  é o correspondente ao maior valor da função log-verossimilhança condicional.

Adotamos que a quantidade de termos autorregressivos e de médias móveis presentes na função log-verossimilhança perfilada são obtidos a partir do modelo NG-ARMA e dos critérios AIC e BIC.

O cálculo do intervalo de confiança para o parâmetro  $\lambda$  é obtido através da inversão da estatística da razão de verossimilhança (RV)  $w = 2\{l(\hat{\lambda}) - l(\lambda^0)\}$ , sendo que  $\lambda^0$  é um valor hipotético. A estatística  $w$  da RV sob a hipótese nula tem distribuição assintótica  $\chi_1^2$ . Portanto o intervalo de confiança  $(100-\alpha)\%$  aproximado é dado por

$$\{\lambda | l(\lambda) > l(\hat{\lambda}) - \frac{1}{2}\chi_1^2(\alpha)\}, \quad (3.10)$$

para mais detalhes sobre intervalos de confiança obtidos a partir de testes de hipótese consulte Shao (2003).

### 3.2.2 Previsão

Supondo que temos as observações até o instante  $t$ , chamando *origem* das previsões, o valor  $\hat{Z}_t^{(\lambda)}(h) = \hat{E}(Z_{t+h}^{(\lambda)} | \mathfrak{F}_t)$  será a previsão de *origem*  $t$  e *horizonte*  $h$ , com  $h \geq 1$ . Adotando  $h = 1$  temos que a previsão a um passo a frente é dada por

$$\hat{Z}_t^{(\lambda)}(1) = \sum_{j=1}^p \hat{\phi}_j z_{t+1-j}^{(\lambda)} + \sum_{j=1}^q \hat{\theta}_j (z_{t+1-j}^{(\lambda)} - \hat{\mu}_{t+1-j}), \quad (3.11)$$

sendo que  $\hat{\phi}_j$  e  $\hat{\theta}_i$  para  $j = 1, \dots, p$  e  $i = 1, \dots, q$  são as estimativas de máxima verossimilhança dos respectivos parâmetros e  $\hat{\mu}_t$  é encontrado adotando  $\hat{\mu}_1 = \dots = \hat{\mu}_q = 0$  e substituindo as estimativas dos parâmetros na equação (3.7). Observe que  $\hat{\mu}_{t+1}$  coincide com  $\hat{Z}_t^{(\lambda)}(1)$ .

Nas previsões para  $h > 1$  utilizamos o fato que

$$\widehat{E}[Z_{t+h-j}^{(\lambda)}|\mathfrak{F}_t] = \begin{cases} z_{t+h-j}^{(\lambda)} & \text{se } h \leq j \\ \widehat{Z}_t^{(\lambda)}(h-j) & \text{se } h > j \end{cases} \quad (3.12)$$

$$\widehat{\mu}_{t+h-j} = \begin{cases} \widehat{\mu}_{t+h-j} & \text{se } h \leq j \\ \widehat{Z}_t^{(\lambda)}(h-j) & \text{se } h > j \end{cases} \quad (3.13)$$

com isso a previsão para  $h$  passos a frente é dada por

$$\widehat{Z}_t^{(\lambda)}(h) = \sum_{j=1}^p \widehat{\phi}_j \widehat{E}[Z_{t+h-j}^{(\lambda)}|\mathfrak{F}_t] + \sum_{j=1}^q \widehat{\theta}_j (\widehat{E}[Z_{t+h-j}^{(\lambda)}|\mathfrak{F}_t] - \widehat{\mu}_{t+h-j}). \quad (3.14)$$

Note que até agora fizemos previsão para a série transformada e livre de sazonalidade, mas estamos interessados na previsão da série original. Para que não precisemos fazer aproximações optamos por prever a série original através da mediana, observe que

$$P(Z_{t+h} \leq z_{t+h}^{(M)}|\mathfrak{F}_t) = 0,5,$$

assim  $z_{t+h}^{(M)}$  é o valor da mediana da distribuição  $Z_{t+h}|\mathfrak{F}_t$ , mas conseguimos escrever a série  $Z_t$  como função da série  $X_t$ , disto temos que

$$P\left(\frac{X_{t(r_1, m_1)}^{\widehat{\lambda}} - \widehat{\mu}_{m_1}}{\widehat{\sigma}_{m_1}} \leq z_{t+h}^{(M)}|\mathfrak{F}_t\right) = 0,5,$$

sendo  $t(r, m) + h = t(r_1, m_1)$ . Definindo  $\mu_t^*(h)$  como a previsão para  $X_{t(r, m)+h}$  temos que

$$\mu_t^*(h) = [\widehat{\lambda}(z_{t+h}^{(M)}\widehat{\sigma}_{m_1} + \widehat{\mu}_{m_1}) + 1]^{1/\widehat{\lambda}},$$

como a distribuição normal generalizada é simétrica em torno da média, e a média é igual a mediana, podemos aproximar a mediana pela previsão da série sem sazonalidade, resultando em

$$\mu_t^*(h) = [\widehat{\lambda}(\widehat{\mu}_{t+h}\widehat{\sigma}_{m_1} + \widehat{\mu}_{m_1}) + 1]^{1/\widehat{\lambda}},$$

para qualquer  $h \geq 1$ .

Note que as previsões só foram feitas para  $\lambda \neq 0$ , pois para  $\lambda = 0$  temos que a transformação de Box-Cox coincide em adotar que a série temporal possui distribuição LNG, logo, a previsão neste caso já foi apresentada na Seção 3.1.1. Portanto, para qualquer valor do parâmetro  $\lambda$  podemos fazer previsões tanto para a série transformada quanto para a série original.

### 3.2.3 Estudo de Simulação

Para o estudo de simulação a geração da série temporal não levará em conta a periodicidade na média e na variância. A geração da série é feita seguindo os seguintes passos:

1. definir os valores dos parâmetros do modelo e dos valores iniciais para  $\mu_t$  e  $z_t^{(\lambda)}$ ;
2. calcular  $\mu_t = \sum_{j=1}^p \phi_j z_{t-j} + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j})$  e gerar  $z_t$  da distribuição  $NG(\mu_t, \sigma, s)$ , para  $t=r^*+1, \dots$ , com  $r^* = \max(p, q)$ ;
3. calcular  $x_t = (\lambda z_t^\lambda + 1)^{1/\lambda}$ ;
4. repetir os itens 2 e 3 até conseguir uma série com  $K$  valores a mais do que se deseja;
5. desconsiderar os  $K$  primeiros valores da série para conseguir o tamanho da série desejada.

Essa geração leva em consideração que  $\lambda \neq 0$ , caso  $\lambda = 0$  então no item 3 devemos calcular  $x_t = \exp(y_t)$ .

Com o objetivo de verificar as propriedades frequentistas dos estimadores de máxima verossimilhança, repetimos 1000 vezes a geração de séries com diferentes tamanhos (100, 200, 500 e 1000), sendo que para cada repetição obtemos inicialmente o valor de  $\lambda$  e a partir dele encontramos o valor das estimativas pontuais e a matriz de informação observada de Fisher. A partir destes valores, calculamos o vício absoluto médio (VAM), o erro quadrático médio (EQM), a média das estimativas (ME), o intervalo de confiança assintótico com 95% e a distribuição amostral de  $\lambda$ . A probabilidade de cobertura (PC) dos parâmetros é o quociente entre o número de intervalos que contém o verdadeiro valor do parâmetro e o número total de intervalos construídos.

Neste estudo de simulação adotamos o modelo BC-NG-ARMA(1,1) com os valores  $\phi_1 = 0,9$ ;  $\theta_1 = 0,2$  e  $\sigma = 0,3$  e, como sabemos que o comportamento das caudas da distribuição normal generalizada se comporta dependendo do valor de  $s$ , assumimos os valores  $s = 1,5$  e  $s = 3,0$ , contemplando assim caudas mais pesadas e mais leves, respectivamente.

Adotamos que o valor verdadeiro de  $\lambda$  é igual a -0,5; 0 e 0,5. Devido aos resultados serem muito parecidos quando apenas substituimos o valor de  $\lambda$  então apresentamos na Tabela 3.4 os resultados das medidas VAM, EQM, ME, PC apenas quando o verdadeiro valor de  $\lambda$  é igual a 0,5.

Tabela 3.4: Estudo de simulação do modelo NG-ARMA considerando as métricas VAM, ME, EQM e PC quando o  $\lambda$  correto é igual a 0,5

	s=1,5				s=3,0			
	ME	VAM	EQM	PC	ME	VAM	EQM	PC
n=100								
$\phi_1$	0,8827	0,0404	0,0030	0,907	0,8864	0,0394	0,0032	0,931
$\theta_1$	0,2002	0,0813	0,0108	0,887	0,1901	0,0865	0,0124	0,908
$\sigma$	0,3059	0,0500	0,0040	0,905	0,3000	0,0248	0,0010	0,902
$s$	1,6836	0,3632	0,3133	0,945	3,7597	1,1086	1,0959	0,955
n=200								
$\phi_1$	0,8927	0,0256	0,0012	0,947	0,8939	0,0262	0,0012	0,947
$\theta_1$	0,1961	0,0595	0,0055	0,929	0,1973	0,0582	0,0054	0,925
$\sigma$	0,2992	0,0326	0,0017	0,930	0,3006	0,0161	0,0004	0,942
$s$	1,5471	0,1960	0,0670	0,952	3,3054	0,5876	0,6955	0,967
n=500								
$\phi_1$	0,8974	0,0162	0,0004	0,948	0,8972	0,0158	0,0004	0,954
$\theta_1$	0,1982	0,0355	0,0020	0,943	0,1989	0,0351	0,0019	0,946
$\sigma$	0,3006	0,0199	0,0006	0,944	0,3003	0,0105	0,0002	0,945
$s$	1,5242	0,1168	0,0220	0,956	3,0844	0,3052	0,1539	0,961
n=1000								
$\phi_1$	0,8983	0,0120	0,0002	0,935	0,8987	0,0110	0,0002	0,952
$\theta_1$	0,1976	0,0272	0,0011	0,941	0,1978	0,0243	0,0010	0,948
$\sigma$	0,3003	0,0146	0,0003	0,931	0,2997	0,0074	0,0001	0,954
$s$	1,5108	0,0820	0,0109	0,936	3,0324	0,1949	0,0617	0,950

Notamos que, aparentemente, o valor de  $\lambda$  não influencia nas métricas estudadas. Observamos também que a média das estimativas se aproxima do verdadeiro valor do parâmetro, o vício absoluto médio e o erro quadrático médio diminuem com o aumento do tamanho da série. A probabilidade de cobertura se apresenta um pouco abaixo da nominal para séries pequenas ( $n=100$ ), mas se aproxima da nominal quando o tamanho da série é igual ou superior a 200.

Verificamos também que o valor do parâmetro  $s$  influencia nas métricas EQM e VAM do parâmetro  $\sigma$ , sendo que para  $s < 2$  o valor destas métricas são maiores do que quando  $s > 2$ . Para os parâmetros  $\phi_1$  e  $\theta_1$  não observamos influência do parâmetro  $s$  nas métricas estudadas.

Os histogramas das distribuições amostrais de  $\hat{\lambda}$  quando o verdadeiro valor de  $\lambda$  é 0,5 e o valor de  $s$  é 1,5 e 3,0 são apresentados nas Figuras 3.5 e 3.6.

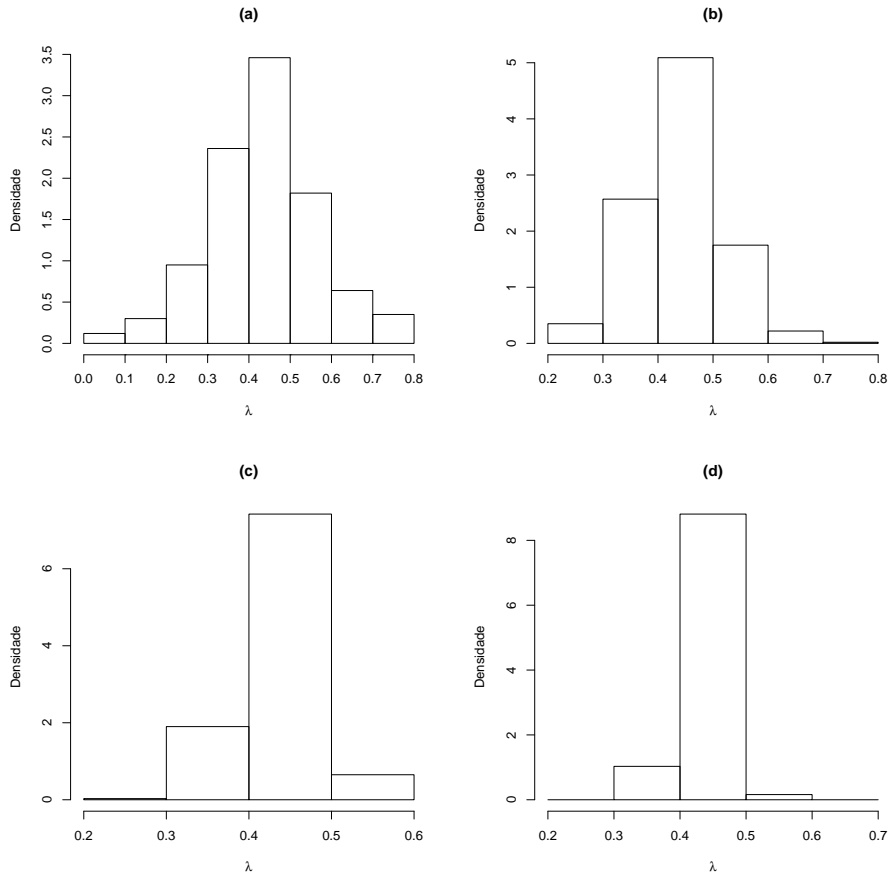


Figura 3.5: Histogramas da distribuição amostral de  $\hat{\lambda}$  quando ajustamos o modelo BC-NG-ARMA sendo  $\lambda = 0,5$  e  $s = 1,5$  os valores verdadeiros e adotando os tamanhos de séries iguais a 100, 200, 500 e 1000 em (a), (b), (c) e (d), respectivamente

Notamos que para séries de tamanho pequeno, a distribuição amostral do parâmetro é mais vaga em comparação à séries com tamanho maior. Verificamos também que o valor verdadeiro de  $\lambda$  é o que apresenta maior probabilidade de ocorrer, para todos os tamanhos de séries estudados. Este comportamento é verificado em todos os cenários estudados.

Na comparação entre as distribuições amostrais obtidas quando utilizamos  $s = 1,5$  e  $s = 3,0$ , observamos que quando adotamos  $s = 3,0$  a distribuição amostral do  $\lambda$  parece ser mais simétrica do que quando usamos  $s = 1,5$  para a geração das séries.

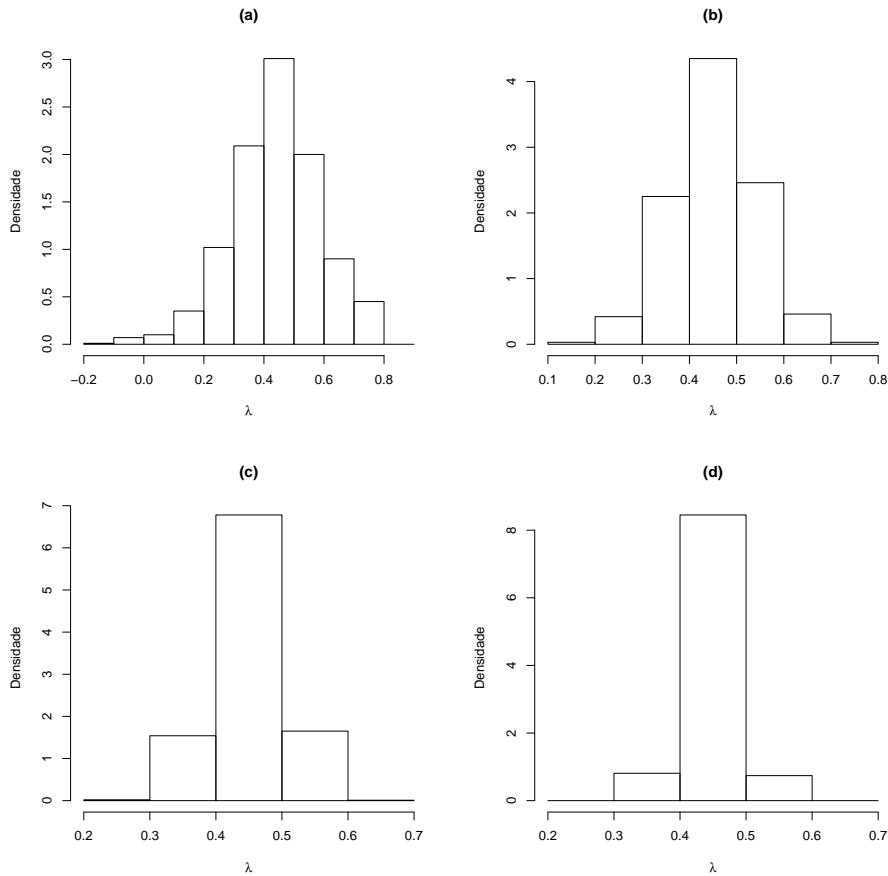


Figura 3.6: Histogramas da distribuição amostral de  $\hat{\lambda}$  quando ajustamos o modelo BC-NG-ARMA sendo  $\lambda = 0,5$  e  $s = 3,0$  os valores verdadeiros e adotando os tamanhos de séries iguais a 100, 200, 500 e 1000 em (a), (b), (c) e (d), respectivamente

### 3.2.4 Aplicação à Série de Furnas

Considerando a série de vazões médias mensais de Furnas observada de 01/1931 a 12/2012, apresentada na Seção 2.3.1, observamos pelo histograma, mostrado na Figura 3.7, que a distribuição da série apresenta assimetria e como já foi exibido na Figura 2.3 a série também possui sazonalidade. Devido a essas características temos que o modelo BC-NG-ARMA se apresenta com uma alternativa para modelar a série em estudo.

Para estimarmos o valor de  $\lambda$  adotamos a ordem (1,1), essa é a ordem que apresentou os melhores valores dos critérios AIC e BIC utilizando as distribuições normal generalizada e normal, respectivamente, como apresentado na Tabela 2.5.

Para o ajuste utilizamos o processo de transformação, eliminação da sazonalidade e otimização da função log-verossimilhança para cada valor de  $\lambda$  dentro do conjunto

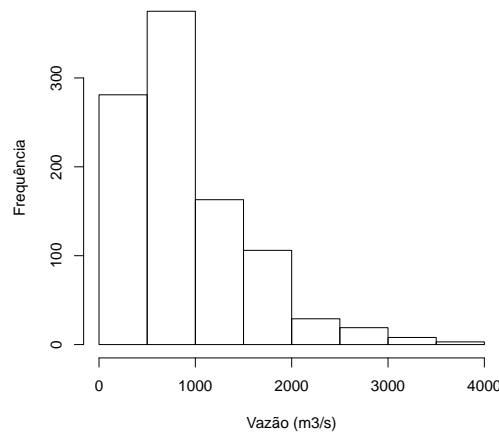


Figura 3.7: Histograma da série de vazões médias mensais de Furnas

$\{-0,9; -0,8; \dots; -0,1; 0; 0,1; \dots; 0,8; 0,9\}$ . Na Figura 3.8 observamos os gráficos com os valores de  $\lambda$  versus o valor do máximo da função log-verossimilhança. Notamos que o maior valor da log-verossimilhança encontrado, quando adotamos a distribuição normal generalizada, é correspondente ao  $\lambda = -0.1$ , e quando adotamos a distribuição normal, o valor é correspondente ao  $\lambda = 0$ .

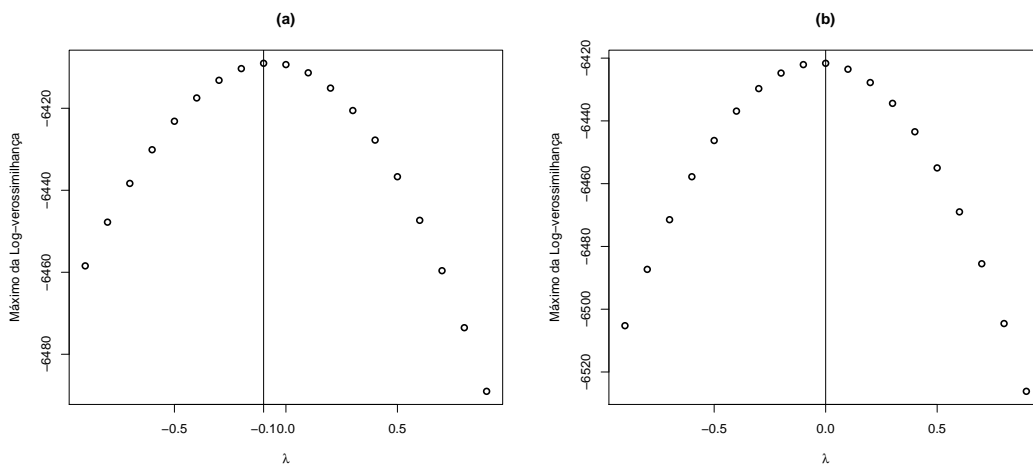


Figura 3.8: Os valores do máximo da função log-verossimilhança para alguns valores de  $\lambda$  quando adotamos os modelos BC-NG-ARMA e BC-N-ARMA, respectivamente em (a) e (b)

Os intervalos de confiança para  $\lambda$  segundo cada um dos modelos são os conjuntos  $\{-0,2; -0,1; 0\}$  e  $\{-0,1; 0; 0,1\}$ , respectivamente quando adotamos as distribuições normal generalizada e normal. Observemos que a intersecção dos intervalos é o conjunto

$\{-0, 1; 0\}$ . Escolhemos para o ajuste o valor de  $\lambda = 0$ , que é a estimativa pontual obtida através do modelo BC-N-ARMA e que está na intersecção dos intervalos de confiança. Assim as estimativas de máxima verossimilhança dos modelos com distribuições normal e normal generalizada e as previsões para a série original são as apresentadas na Seção 3.1.2.

Notamos então que mesmo escolhendo a estimativa de  $\lambda$  oriunda do ajuste com a distribuição normal, o modelo BC-NG-ARMA se ajustou melhor à série em estudo em comparação ao modelo BC-N-ARMA, resultado obtido analisando os valores dos critérios AIC e BIC apresentados na Tabela 3.1.

### 3.2.5 Aplicação à Série de Sobradinho

Adotamos a série de vazões médias mensais afluentes da usina hidrelétrica de Sobradinho observada de 01/1931 a 12/1998, apresentada na Figura 3.9(a). Para que mais a frente possamos comparar as previsões com os valores reais desconsideramos do ajuste os últimos 12 meses. A distribuição da série em questão exibe assimetria, como pode ser visto no histograma apresentado na Figura 3.9(b).

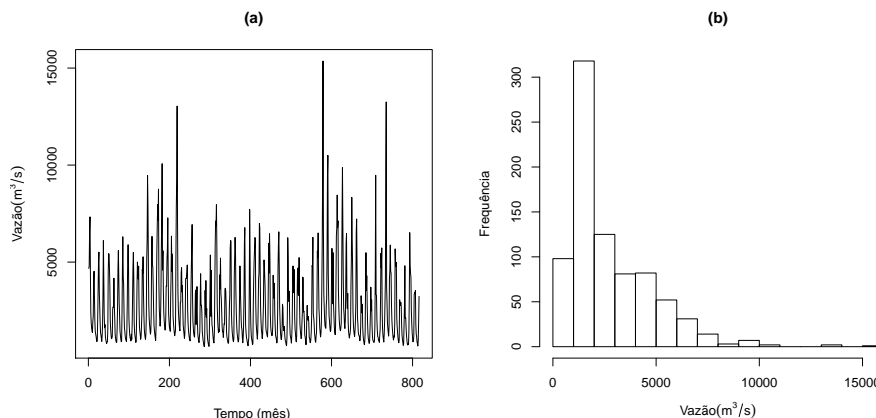


Figura 3.9: Série de vazões médias mensais da usina hidrelétrica de Sobradinho em (a) e seu histograma em (b)

Para encontrar o valor de  $\lambda$  precisamos inicialmente encontrar a ordem que melhor representa a série temporal de Sobradinho utilizando os modelos NG-ARMA e ARMA com distribuição normal, para os ajustes eliminamos a sazonalidade adotando a equação (2.4). São mostrados na Tabela 3.5 os valores dos critérios AIC e BIC de diferentes ordens.



Tabela 3.5: Seleção da melhor ordem a partir dos modelos NG-ARMA e N-ARMA utilizando os critérios AIC e BIC

Modelo	Log-Verossimilhança	AIC	BIC
NG-ARMA(1,0)	-747,47	1500,93	1515,00
NG-ARMA(2,0)	-742,22	1492,44	1511,19
NG-ARMA(0,1)	-924,74	1855,49	1869,56
NG-ARMA(1,1)	<b>-740,24</b>	<b>1488,47</b>	<b>1507,23</b>
ARMA(1,0)	-806,60	1617,20	1626,58
ARMA(2,0)	-805,18	1616,35	1630,42
ARMA(0,1)	-933,56	1871,43	1738,50
ARMA(1,1)	-805,50	1617,00	1631,06

Observamos através da Tabela 3.5 que o melhor modelo, entre os apresentados, é o NG-ARMA(1,1). Observamos também que entre aqueles que utilizam a distribuição normal, o melhor é o ARMA(1,0), para o critério BIC, enquanto que para o critério AIC é o ARMA(2,0). Para a estimação de  $\lambda$  optamos por adotar os modelos BC-NG-ARMA(1,1) e BC-N-ARMA(1,0).

Para estimar o valor de  $\lambda$  utilizamos o processo de transformação, eliminação da sazonalidade e otimização da função log-verossimilhança para cada valor de  $\lambda$  dentro do conjunto  $\{-0,9; -0,8; \dots; -0,1; 0; 0,1; \dots; 0,8; 0,9\}$ . Na Figura 3.10 observamos que o valor de  $\lambda$  correspondente ao maior valor da função log-verossimilhança do modelo BC-NG-ARMA(1,1) é igual a -0,3, e quando adotamos o modelo BC-N-ARMA(1,0) o valor é de -0,1.

Os intervalos de confiança para  $\lambda$  são os conjuntos:  $\{-0,4; -0,3; -0,2; -0,1\}$  e  $\{-0,2; -0,1; 0\}$ , respectivamente para os modelos BC-NG-ARMA e BC-N-ARMA. Notemos que a intersecção é o conjunto  $\{-0,2; -0,1\}$ . Escolhemos novamente para o ajuste o valor de  $\lambda$  que mais favorece o modelo com a distribuição normal, ou seja, adotamos o valor de  $\lambda = -0,1$ .

A partir da série transformada, considerando  $\lambda = -0,1$ , e livre de sazonalidade, adotando a da equação (3.6), construímos os gráficos das funções de autocorrelação e de autocorrelação parcial que são apresentados na Figura 3.11. Notamos na Figura 3.11(a)

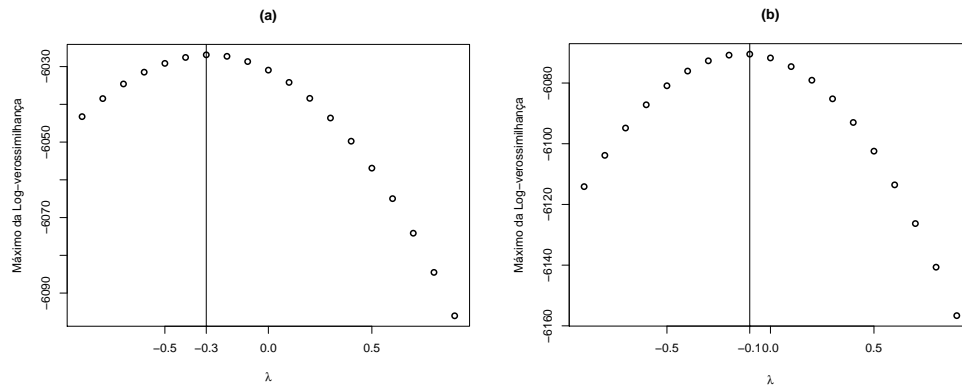


Figura 3.10: Os valores do máximo da função log-verossimilhança para alguns valores de  $\lambda$  considerando os modelos BC-NG-ARMA e BC-N-ARMA, respectivamente em (a) e (b)

que existe uma forte correlação nos primeiros lags, já na Figura 3.11(b) a função de autocorrelação parcial apresenta correlação mais forte nos lags iguais a 1, 4, 11 e 13.

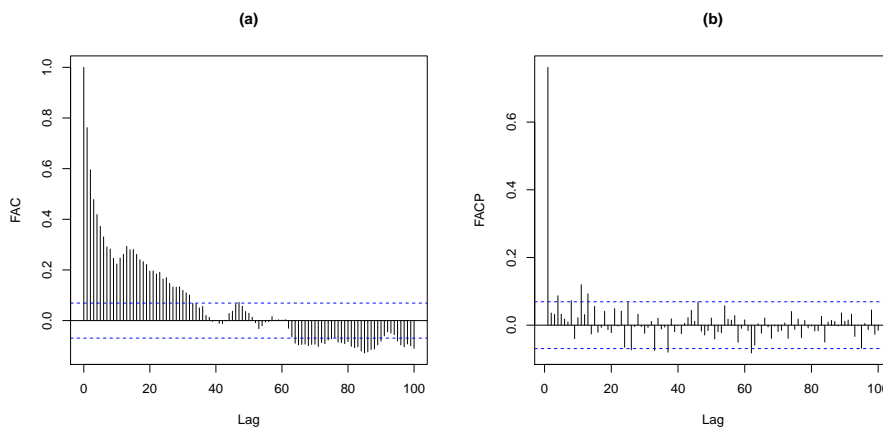


Figura 3.11: Funções de autocorrelação e autocorrelação parcial da série de Sobradinho livre de sazonalidade em (a) e em (b), respectivamente

Ajustamos diversos modelos usando as distribuições normal e normal generalizada, com diferentes ordens, a escolha do modelo que melhor representa a série é feita utilizando os critérios AIC e BIC. O resultados dos critérios são apresentados na Tabela 3.6.

Observando os resultados apresentados na Tabela 3.6 notamos que o melhor modelo segundo os critérios é o BC-NG-ARMA(3,0), mas a estimativa do segundo termo autorregressivo é não significativa, assim adotamos que o modelo que melhor se ajusta à série é o BC-NG-ARMA(1,1), pois apresenta o menor valor dos critérios quando elimina-

Tabela 3.6: Seleção entre os modelos BC-NG-ARMA e BC-N-ARMA utilizando os critérios AIC e BIC para explicar à série de Furnas

	AIC	BIC
BC-N-ARMA(1,0)	12144,96	12154,34
BC-N-ARMA(2,0)	12127,29	12141,36
BC-N-ARMA(1,1)	12145,84	12159,91
BC-N-ARMA(0,1)	12421,49	12430,87
BC-NG-ARMA(1,0)	12069,92	12083,99
BC-NG-ARMA(2,0)	12047,68	12066,44
BC-NG-ARMA(3,0)	<b>12020,72</b>	<b>12044,17</b>
BC-NG-ARMA(1,1)	12046,61	12065,36
BC-NG-ARMA(0,1)	12423,23	12437,30

mos da comparação o modelo BC-NG-ARMA(3,0).

Na Tabela 3.7 observamos os valores das estimativas de máxima verossimilhança e os intervalos de confiança assintóticos obtidos no ajuste do modelo BC-NG-ARMA(1,1). Notamos que o valor de  $s$  é menor que 2, implicando que as caudas da distribuição da série são mais pesadas do que as da distribuição normal. Constatamos ainda que todos os parâmetros são significativos.

Tabela 3.7: Estimativas pontual e intervalar dos parâmetros do modelo BC-NG-ARMA(1,1) que melhor se ajusta à série de Sobradinho

Parâmetro	EMV	IC
$\phi_1$	0,8830	(0,8479; 0,9180)
$\theta_1$	-0,0869	(-0,1210; -0,0529)
$\sigma$	0,4753	(0,3780; 0,5725)
$s$	1,0169	(0,8756; 1,1581)

A partir do ajuste do modelo BC-NG-ARMA(1,1) com  $\lambda = -0,1$ , calculamos o resíduo, que é a diferença entre o valor da série transformada e livre de sazonalidade e a média estimada pelo modelo, para que possamos analisar o ajuste. Na Figura 3.12 mostramos as funções de autocorrelação e autocorrelação parcial dos resíduos. Notamos que a correlação não desaparece nos lags 2, 3, 10 e 13, e que a partir da função de

autocorrelação parcial observamos diversos valores significativos. Comparando o ajuste das ordens (1,1) e (3,0), quando ajustamos o modelo com a ordem (3,0) à série observamos que além do segundo parâmetro autorregressivo não ser significativo os resíduos apresenta correlação nos lags 3, 10 e 13, assim optamos utilizar a ordem (1,1).

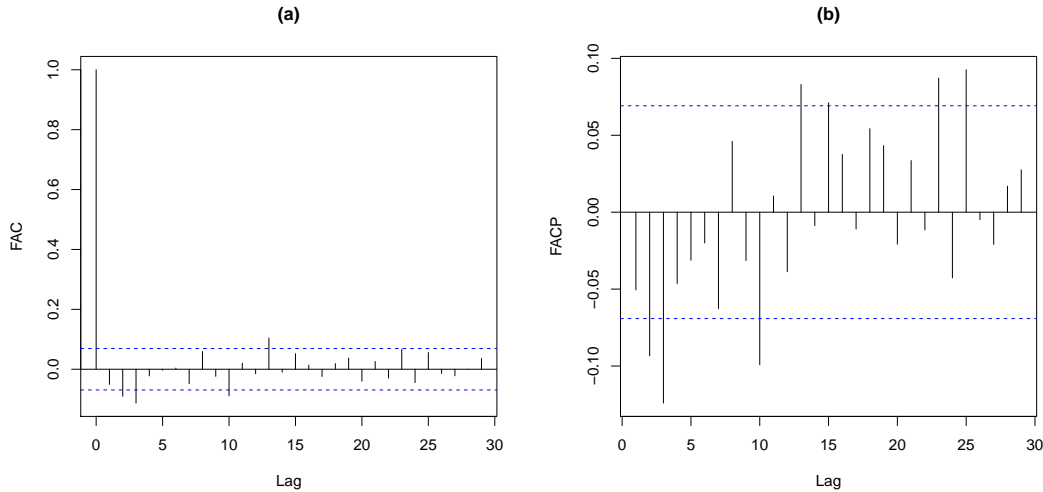


Figura 3.12: Funções de autocorrelação e autocorrelação parcial dos resíduos do ajuste do modelo BC-NG-ARMA(1,1) com  $\hat{\lambda} = -0,1$  à série de Sobradinho em (a) e em (b), respectivamente

A partir do gráfico quantil teórico da distribuição normal generalizada versus quantil observado dos resíduos, apresentado na Figura 3.13, verificamos que a distribuição proposta apresentou um bom ajuste aos dados.

Na Figura 3.14 mostramos os valores preditos nos dois cenários para a série real, sendo que os cenários são os mesmos já definidos anteriormente. Observamos claramente que os valores previstos para o 2º cenário se aproximam mais dos verdadeiros valores do que os valores previstos no 1º cenário. Notamos também que nos meses em que a vazão apresenta um maior volume, as previsões tendem a cometer um maior erro comparado com os meses de vazão menor.

Podemos quantificar a qualidade dos dois cenários de previsões através das medidas MAPE, EAM e EQM, que estão apresentadas na Tabela 3.8. Observamos que a qualidade dos valores preditos no 2º cenário é amplamente superior em comparação aos obtidos no 1º cenário.

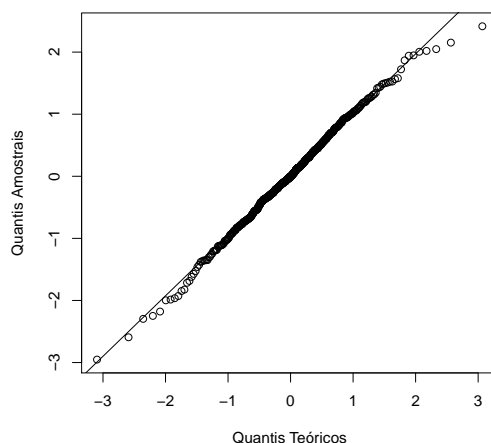


Figura 3.13: Gráfico quantil teórico da distribuição normal generalizada versus quantil observado dos resíduos do ajuste do modelo BC-NG-ARMA(1,1) com  $\hat{\lambda} = -0,1$  à série de Sobradinho

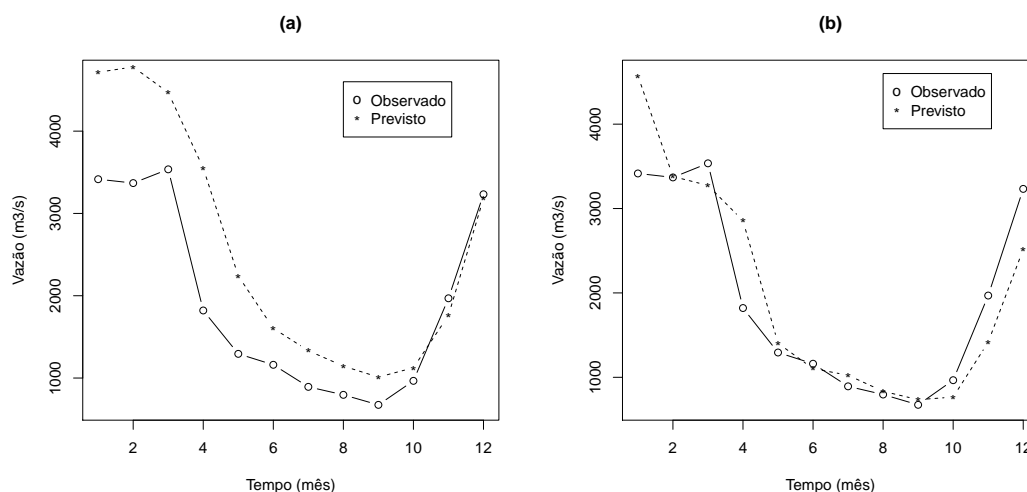


Figura 3.14: Previsões para a série de Sobradinho adotando o modelo BC-NG-ARMA(1,1) com  $\hat{\lambda} = -0,1$  e considerando o 1º cenário em (a), e o 2º cenário em (b)

Tabela 3.8: Avaliação das previsões para a série de Sobradinho feitas a partir do modelo BC-NG-ARMA(1,1) com  $\lambda = -0,1$  segundo os critérios MAPE, EQM e EAM

Cenário	MAPE(%)	EQM	EAM
1º	39,6	702.413,1	666,8
2º	17,7	281.210,9	362,3

### 3.3 Conclusão

Neste capítulo propusemos uma nova classe de modelos para séries temporais que apresentam assimetria e que a transformação de Box-Cox consegue transformá-las em séries aproximadamente simétricas e com variância constante. Para esta nova classe de modelos apresentamos as funções de verossimilhança e log-verossimilhança e um método para o cálculo de previsões. A partir do estudo de simulação observamos ótimas propriedades das estimativas de máxima verossimilhança.

Notamos a aplicabilidade do modelo proposto através dos ajustes do mesmo às séries de vazões médias mensais das usinas hidrelétricas de Furnas e de Sobradinho, sendo que os ajustes conseguiram capturar a correlação na maioria dos lags e as caudas mais pesadas da distribuição dos dados. Também foi observado em ambas as séries que o método de previsão fornece resultado satisfatório.

# Capítulo 4

## Modelo Normal Generalizada Autorregressivo Periódico

Modelos autorregressivos periódicos são extensões dos modelos ARMA no sentido de flexibilizar a estrutura de correlação, possibilitando uma melhor capacidade para se ajustar a séries que apresentam correlação periódica, resultando em ajustes mais precisos e consequentemente em previsões mais apuradas.

Motivados por séries que apresentam periodicidade tanto na média e variância quanto na função de correlação, neste capítulo apresentamos um modelo denominado por normal generalizada autorregressivo periódico (NG-PAR), que modela a média da série temporal. Na Seção 4.1 formulamos o modelo NG-PAR e a função de máxima verossimilhança é apresentada na Seção 4.2. Na Seção 4.3 mostramos um método para o cálculo de previsões. Estudos de simulações para verificar propriedades frequentistas dos estimadores de máxima verossimilhança e o comportamento dos critérios AIC e BIC são apresentados na Seção 4.4. Comparamos o desempenho do modelo NG-PAR com o modelo PAR com distribuição normal utilizando a série de vazões da usina hidrelétrica de Furnas na Seção 4.5. Alguns comentários são apresentados na Seção 4.6.

## 4.1 Modelo Normal Generalizada Autorregressivo Periódico

Considere  $\{X_{t(r,m)}; r = 1, \dots, R; m = 1, \dots, S\}$  uma série temporal periódica. Eliminamos a sazonalidade presente adotando a equação dada por

$$Z_{t(r,m)} = \frac{X_{t(r,m)} - \mu_m}{\sigma_m}, \quad (4.1)$$

sendo que  $\mu_m = E\{X_{t(r,m)}\}$ ,  $\sigma_m = \sqrt{Var\{X_{t(r,m)}\}}$  e  $\{Z_{t(r,m)}; r = 1, \dots, R; m = 1, \dots, S\}$  é uma série temporal que apresenta função de correlação periódica. Comparando as equações (2.4) e (4.1) observamos que a diferença está no índice de tempo da série  $Z_t$ , ou seja, no Capítulo 2 não fizemos nenhuma suposição acerca da série  $Z_t$ , sendo que neste capítulo estamos impondo que a série  $Z_t$  apresente uma estrutura de correlação periódica. Por isso optamos por denotar o índice de tempo da série  $Z_t$  por uma função de  $r$  e  $m$ , pois agora temos interesse em saber quem são os índices  $r$  e  $m$ .

Assumimos que a distribuição condicional da série  $Z_{t(r,m)}$  para cada  $m$ ,  $m = 1, \dots, S$ , dada toda a informação passada do processo possui função densidade de probabilidade normal generalizada com parâmetros  $\mu_m^*$ ,  $\sigma_m^*$  e  $s_m$ , ou seja,

$$Z_{t(r,m)} | \mathfrak{F}_{t(r,m)-1} \sim NG(\mu_m^*, \sigma_m^*, s_m),$$

sendo que  $\mathfrak{F}_{t(r,m)-1}$  é o conjunto de informações das observações até o instante  $t(r, m) - 1$ , definido como

$$\mathfrak{F}_{t(r,m)-1} = \{z_1, \dots, z_{t(r,m)-1}\}, \quad (4.2)$$

e  $\mu_m^* = E[Z_{t(r,m)} | \mathfrak{F}_{t(r,m)-1}]$  sendo dado por

$$\mu_m^* = \sum_{j=1}^{p_m} \phi_{j,m} z_{t(r,m)-j}, \quad (4.3)$$

em que  $\mu_m^*$  é a média da distribuição normal generalizada da  $m$ -ésima observação dentro de qualquer período e  $\phi_{1,m}, \dots, \phi_{p_m,m}$  são os coeficientes do modelo autorregressivo para a  $m$ -ésima observação.



## 4.2 Estimação de Máxima Verossimilhança

Denotemos os parâmetros do modelo a serem estimados por  $(\phi_1, \dots, \phi_S, \sigma, \mathbf{s})$  sendo que  $\phi_m = \{\phi_{1,m}, \dots, \phi_{p_m,m}\}$  com  $m = 1, \dots, S$ ,  $\sigma = \{\sigma_1^*, \dots, \sigma_S^*\}$  e  $\mathbf{s} = \{s_1, \dots, s_S\}$ . Seja  $\mathbf{z} = \{z_1, \dots, z_n\}$  a série temporal observada,  $\mu_t^*$  como descrito na equação (4.3) e  $r_{1,m} = 1$  se  $p_m < m$  ou  $r_{1,m} = 2$  se  $m \leq p_m < S$ . Para cada  $m$  as estimativas dos parâmetros podem ser obtidas utilizando a função verossimilhança condicional introduzida por Cox (1975), que é dada por

$$\begin{aligned}
L(\phi_m, \sigma_m^*, s_m | \mathbf{z}) &= \prod_{r=r_{1,m}}^R f(z_{t(r,m)} | \mathfrak{F}_{t(r,m)-1}) \\
&= \prod_{r=r_{1,m}}^R \frac{s}{2\sigma_m^* \Gamma(1/s)} \exp \left\{ - \left| \frac{z_{t(r,m)} - \mu_{t(r,m)}^*}{\sigma_m^*} \right|^s \right\} \\
&= \left( \frac{s}{2\sigma_m^* \Gamma(1/s)} \right)^{R-r_{1,m}} \exp \left\{ - \frac{1}{(\sigma_m^*)^s} \sum_{r=r_{1,m}}^R |z_{t(r,m)} - \mu_{t(r,m)}^*|^s \right\} \\
&= \left( \frac{s}{2\sigma_m^* \Gamma(1/s)} \right)^{R-r_{1,m}} \times \\
&\quad \exp \left\{ - \frac{1}{(\sigma_m^*)^s} \sum_{r=r_{1,m}}^R \left| z_{t(r,m)} - \sum_{j=1}^{p_m} \phi_{j,m} z_{t(r,m)-j} \right|^s \right\}, \tag{4.4}
\end{aligned}$$

a função de verossimilhança para todos os parâmetros é dada pelo produto das funções de verossimilhança de cada  $m$ , ou seja,

$$L(\phi_1, \dots, \phi_S, \sigma, \mathbf{s} | \mathbf{z}) = \prod_{m=1}^S L(\phi_m, \sigma_m^*, s_m | \mathbf{z}). \tag{4.5}$$

Observamos pela equação (4.5) que a função de log-verossimilhança é dada por

$$\begin{aligned}
l(\phi_1, \dots, \phi_S, \sigma, \mathbf{s} | \mathbf{z}) &= \log L(\phi_1, \dots, \phi_S, \sigma, \mathbf{s} | \mathbf{z}) \\
&= \sum_{m=1}^S \log L(\phi_m, \sigma_m^*, s_m | \mathbf{z}) \\
&= \sum_{m=1}^S l(\phi_m, \sigma_m^*, s_m | \mathbf{z}),
\end{aligned}$$

sendo que

$$\begin{aligned}
l(\phi_m, \sigma_m^*, s_m | \mathbf{z}) &= (R - r_{1,m}) \log \left( \frac{s}{2\sigma_m^* \Gamma(1/s)} \right) \\
&\quad - \frac{1}{(\sigma_m^*)^s} \sum_{r=r_{1,m}}^R \left| z_{t(r,m)} - \sum_{j=1}^{p_i} \phi_{j,i} z_{t(r,m)-j} \right|^s, \tag{4.6}
\end{aligned}$$

ou seja, as estimativas de mxima verossimilhana dos parmetros para cada  $m$  podem ser encontradas de maneira separada. Mas observamos pela equaco (4.4) que a funo de verossimilhana   bastante complexa, por isso indicamos m todos iterativos para a obteno das estimativas de mxima verossimilhana, sendo que nesse caso a funo que   bastante utilizada   a log-verossimilhana, que   dada por (4.6). Os intervalos de confiana so obtidos assumindo normalidade assint tica dos estimadores de mxima verossimilhana.

Algumas maneiras de escolher as ordens do modelo PAR so discutidas na literatura, por exemplo, para cada per odo  $m$ , a ordem ser  $i$  se todos os coeficientes estimados da funo de autocorrelaco per dica (FACP) maiores que  $i$  sejam no significativos, ou ainda, a ordem  $i$  seja escolhida de tal maneira que todos os coeficientes da FACP menores que  $i$  sejam significativos. Uma extenso   sugerida por Oliveira e Souza (2011) que analisa os dois m todos utilizando bootstrap para a construo dos intervalos de confiana para os coeficientes da FACP.

Neste trabalho iniciamos a investigao da ordem do modelo NG-PAR para cada um dos per odos atrav s das funes de autocorrelaco per dica e autocorrelaco parcial per dica, mas optamos por escolher a ordem atrav s dos crit rios AIC e BIC calculados em cada um dos per odos, que so dados por

$$\begin{aligned} AIC_m &= -2l(\hat{\phi}_{1,m}, \dots, \hat{\phi}_{p_m,m}, \hat{\sigma}_m^*, \hat{s}_m | \mathbf{z}) + 2(p_m + 2) \\ BIC_m &= -2l(\hat{\phi}_{1,m}, \dots, \hat{\phi}_{p_m,m}, \hat{\sigma}_m^*, \hat{s}_m | \mathbf{z}) + (p_m + 2) \log(R - r_{1,m}), \end{aligned}$$

sendo que  $\hat{\phi}_{i,m}, \hat{\sigma}_m, \hat{s}_m$ , com  $i = 1, \dots, p_m$  so as estimativas de mxima verossimilhana dos respectivos parmetros, desde que todos os parmetros autorregressivos sejam significativos. Os valores dos crit rios AIC e BIC resultante do ajuste so dados por

$$\begin{aligned} AIC &= \sum_{m=1}^S AIC_m \\ BIC &= \sum_{m=1}^S BIC_m. \end{aligned}$$

### 4.3 Previsão

Estamos interessados em prever valores da série original, mas para isso inicialmente prevemos valores da série sem sazonalidade e depois adicionamos o efeito da sazonalidade para encontramos previsões para a série original.

Suponha que temos as observações até um instante  $t$ , sendo  $t = (r_0 - 1)S + m_0$ , e queremos agora prever as observações  $h$  passos a frente. Adotamos a previsão com erro quadrático médio mínimo, ou seja, assumimos que o valor previsto é dado através da média, portanto a previsão de *origem*  $t$  e *horizonte*  $h$  é dada por  $\widehat{X}_{(r_0-1)S+m_0}(h) = \widehat{E}(X_{(r_0-1)S+m_0+h}|\mathfrak{F}_t)$ .

Observe que a previsão a um passo a frente ( $h = 1$ ) para a série sem sazonalidade utilizando o modelo NG-PAR é dada por

$$\begin{aligned}\widehat{Z}_{t(r_0, m_0)}(1) &= \widehat{E}(Z_{t(r_0, m_0)+1}|\mathfrak{F}_{t(r_0, m_0)}) \\ &= \sum_{j=1}^{p_{m_0+1}} \widehat{\phi}_{j, m_0+1} z_{t(r_0, m_0)+1-j},\end{aligned}\quad (4.7)$$

sendo que  $\widehat{\phi}_{j, m_0+1}$  com  $j = 1, \dots, p_{m_0+1}$  são as estimativas de máxima verossimilhança dos respectivos parâmetros. Observando as equações (4.3) e (4.7) notamos que a equação de previsão a um passo a frente é dada através da equação da média da distribuição normal generalizada da observação  $m_0 + 1$  dentro de qualquer período, substituindo os parâmetros  $\phi_{j, m_0+1}$  por suas estimativas de máxima verossimilhança.

Observe que para prever o valor no instante  $t+2$  precisamos do valor do instante  $t+1$ . Para resolver esse problema podemos ou observar o valor no instante  $t+1$  e novamente fazer a previsão a um passo a frente ou considerar o valor predito no instante  $t+1$  como sendo o valor observado e assim prever o instante  $t+2$ .

Quando não há o acréscimo de informações, a previsão com *horizonte* maior que 1 ( $h > 1$ ) precisa de todas as previsões até o instante  $t+h-1$  para conseguir prever o valor da série no instante  $t+h$ , por isso definimos que

$$\widehat{E}[Z_{t(r_0, m_0)+h-j}|\mathfrak{F}_{t(r_0, m_0)}] = \begin{cases} z_{t(r_0, m_0)+h-j} & \text{se } h \leq j \\ \widehat{Z}_{t(r_0, m_0)}(h-j) & \text{se } h > j, \end{cases}\quad (4.8)$$

com isso a previsão para qualquer  $h \geq 1$  é dada por

$$\begin{aligned}\widehat{Z}_{t(r_0, m_0)}(h) &= \widehat{E}(Z_{t(r_0, m_0)+h} | \mathfrak{F}_t) \\ &= \widehat{E}(Z_{t(r_1, m_1)} | \mathfrak{F}_t) \\ &= \sum_{j=1}^{p_{m_1}} \widehat{\phi}_{j, m_1} \widehat{E}[Z_{t(r_1, m_1)-j} | \mathfrak{F}_t],\end{aligned}\tag{4.9}$$

em que  $(r_0 - 1)S + m_0 + h = (r_1 - 1)S + m_1$  e  $\widehat{\phi}_{j, m_1}$  com  $j = 1, \dots, p_{m_1}$  são as estimativas de máxima verossimilhança dos respectivos parâmetros. Observe que com as equações (4.7) e (4.9) conseguimos prever a série sem sazonalidade para qualquer *horizonte*. Mas o nosso interesse é a obtenção das previsões para a série original, sendo assim note que

$$\begin{aligned}\widehat{E}(X_{t(r_0, m_0)+h} | \mathfrak{F}_t) &= \widehat{E}(X_{t(r_1, m_1)} | \mathfrak{F}_t) \\ &= \widehat{E}(\widehat{\mu}_{m_1} + \widehat{\sigma}_{m_1} Z_{t(r_1, m_1)} | \mathfrak{F}_t) \\ &= \widehat{\mu}_{m_1} + \widehat{\sigma}_{m_1} \widehat{E}(Z_{t(r_1, m_1)} | \mathfrak{F}_t) \\ &= \widehat{\mu}_{m_1} + \widehat{\sigma}_{m_1} \widehat{Z}_{t(r_1, m_1)}(h),\end{aligned}\tag{4.10}$$

ou seja, a partir da equação (4.10) temos que a previsão para a série original depende da previsão para a série sem sazonalidade e das estimativas da média e do desvio padrão da  $m$ -ésima observação de todos os períodos, para  $m = 1, \dots, S$ .

## 4.4 Estudo de Simulação

Antes de qualquer estudo de simulação precisamos primeiro esquematizar o processo de geração de séries temporais que seguem o modelo NG-PAR. Apresentamos os passos utilizados para a geração a seguir

1. definir os valores dos parâmetros  $S, R, p_1, \dots, p_S, \phi_0, \dots, \phi_m, \sigma, s$ ;
2. definir os valores iniciais de  $z_{t(r, m)}$ ;
3. calcular  $\mu_{t(r, m)}^* = \sum_{j=1}^{p_m} \phi_{j, m} z_{t(r, m)-j}$ ,
4. gerar  $z_{t(r, m)}$  da distribuição  $\text{NG}(\mu_{t(r, m)}^*, \sigma_m^*, s_m)$ ,
5. repetir os itens 3 e 4 até conseguir uma série com  $K$  valores a mais do que se deseja;

6. desconsiderar os  $K$  primeiros valores da série para conseguir o tamanho da série desejada.

Para que possamos analisar algumas propriedades frequentistas dos estimadores de máxima verossimilhança, repetimos 1000 vezes o processo de geração e estimação, considerando diferentes tamanhos de séries, sempre considerando  $S = 12$  e adotando  $R = 30, 50, 80$  e  $100$ , com isso estamos considerando séries com tamanho igual a  $360, 600, 960$  e  $1200$ , sendo que para cada repetição no processo de estimação o interesse é a obtenção das estimativas de máxima verossimilhança e a matriz de informação de Fisher. A partir destes valores, calculamos o vício absoluto médio (VAM), o erro quadrático médio (EQM), a média das estimativas (ME) e a probabilidade de cobertura (PC) dos parâmetros.

Neste primeiro estudo de simulação os valores adotados para os parâmetros são apresentados na Tabela 4.1. Observe que no vetor  $\mathbf{s}$  adotamos tanto valores maiores quanto valores menores que 2, isso para que possamos contemplar caudas mais leves e mais pesadas do que as da distribuição normal. Os resultados das métricas ME, VAM, EQM e PC são apresentados nas Tabelas 4.2, 4.3, 4.4 e 4.5.

Tabela 4.1: Parâmetros adotados para o estudo de simulação do modelo NG-PAR para a análise das métricas ME, VAM, EQM e PC

	m											
	1	2	3	4	5	6	7	8	9	10	11	12
$\phi_{1,m}$	0,60	0,50	0,40	0,35	0,30	0,70	0,80	0,75	0,60	0,50	0,30	0,55
$\sigma_m^*$	1,10	1,50	0,90	0,85	1,20	1,00	1,80	1,30	0,60	0,50	1,40	0,75
$s_m$	1,60	1,50	3,00	1,70	2,00	1,40	1,80	2,10	2,50	3,20	1,60	3,00

Através dos resultados apresentados nas Tabelas 4.2, 4.3, 4.4 e 4.5 notamos que a média das estimativas se apresentam próximas dos verdadeiros valores dos parâmetros mesmo com séries de tamanho pequeno, exceto para  $s_m$ . Observamos também que o vício absoluto médio e o erro quadrático médio diminuem com o incremento do tamanho da série. A probabilidade de cobertura para tamanho de série igual a 360 se apresenta, em geral, abaixo da nominal, mas com o aumento do tamanho da série a probabilidade de cobertura tende à nominal, contudo é observado uma baixa probabilidade de cobertura para o parâmetro  $\phi_{1,m}$  para  $m = 2$  e  $6$ .

Tabela 4.2: Estudo de simulação do modelo NG-PAR considerando as métricas VAM, ME, EQM e PC para  $m = 1, 2$  e  $3$ 

	m=1				m=2				m=3			
	ME	VAM	EQM	PC	ME	VAM	EQM	PC	ME	VAM	EQM	PC
	R=30				R=30				R=30			
$\phi_{1,m}$	0,5965	0,1849	0,0563	0,7620	0,5000	0,1974	0,0647	0,7290	0,3987	0,0606	0,0060	0,8990
$\sigma_m^*$	1,0779	0,2963	0,1373	0,8740	1,4826	0,4184	0,2754	0,8840	0,8465	0,1360	0,0325	0,9130
$s_m$	2,2254	0,9944	3,5255	0,9200	2,0170	0,8547	2,7410	0,9070	3,8660	1,6508	7,4639	0,9170
	R=50				R=50				R=50			
$\phi_{1,m}$	0,6044	0,1352	0,0297	0,8400	0,5104	0,1474	0,0350	0,8100	0,3983	0,0460	0,0035	0,9220
$\sigma_m^*$	1,1020	0,2168	0,0730	0,9160	1,5052	0,3214	0,1608	0,9140	0,8902	0,0944	0,0145	0,9290
$s_m$	1,9372	0,6099	1,6181	0,9480	1,7750	0,5325	0,9492	0,9530	3,9911	1,4950	6,1332	0,9580
	R=80				R=80				R=80			
$\phi_{1,m}$	0,5939	0,1081	0,0183	0,9030	0,5060	0,1146	0,0209	0,8680	0,4007	0,0349	0,0019	0,9340
$\sigma_m^*$	1,1100	0,1697	0,0453	0,9270	1,4908	0,2578	0,1066	0,9150	0,8919	0,0744	0,0093	0,9320
$s_m$	1,7697	0,3967	0,3139	0,9490	1,6305	0,3636	0,3176	0,9270	3,5508	1,0110	2,5935	0,9560
	R=100				R=100				R=100			
$\phi_{1,m}$	0,5959	0,0933	0,0138	0,9080	0,5026	0,1033	0,0168	0,8850	0,3998	0,0311	0,0016	0,9220
$\sigma_m^*$	1,1027	0,1553	0,0371	0,9270	1,5001	0,2272	0,0789	0,9290	0,8978	0,0705	0,0080	0,9280
$s_m$	1,7196	0,3298	0,2040	0,9550	1,5906	0,2932	0,1648	0,9560	3,4967	0,9218	2,5567	0,9420

Tabela 4.3: Estudo de simulação do modelo NG-PAR considerando as métricas VAM, ME, EQM e PC para  $m = 4, 5$  e  $6$ 

	m=4				m=5				m=6			
	ME	VAM	EQM	PC	ME	VAM	EQM	PC	ME	VAM	EQM	PC
	R=30				R=30				R=30			
$\phi_{1,m}$	0,3448	0,1348	0,0296	0,8220	0,2881	0,1348	0,0567	0,8530	0,6953	0,1593	0,0417	0,6990
$\sigma_m^*$	0,8457	0,2066	0,0673	0,8950	1,1748	0,2066	0,0970	0,8970	1,0022	0,3085	0,1429	0,8820
$s_m$	2,3273	0,9897	3,5693	0,9260	2,7117	0,9897	4,2790	0,9320	1,9095	0,8240	2,5316	0,9290
	R=50				R=50				R=50			
$\phi_{1,m}$	0,3533	0,0999	0,0160	0,8860	0,3100	0,0999	0,0359	0,8800	0,6943	0,1254	0,0253	0,7830
$\sigma_m^*$	0,8541	0,1646	0,0423	0,8900	1,2029	0,1646	0,0632	0,9090	1,0174	0,2385	0,0897	0,9040
$s_m$	2,0754	0,6718	1,4267	0,9440	2,5538	0,6718	2,4408	0,9450	1,6852	0,5214	0,9228	0,9560
	R=80				R=80				R=80			
$\phi_{1,m}$	0,3486	0,0763	0,0093	0,9150	0,3015	0,0763	0,0194	0,9270	0,6985	0,0899	0,0133	0,8520
$\sigma_m^*$	0,8563	0,1255	0,0248	0,9200	1,2114	0,1255	0,0347	0,9180	1,0054	0,1849	0,0522	0,9340
$s_m$	1,9189	0,4627	0,6622	0,9470	2,3133	0,4627	0,8794	0,9520	1,5282	0,3256	0,3141	0,9460
	R=100				R=100				n=100			
$\phi_{1,m}$	0,3525	0,0696	0,0076	0,9300	0,3016	0,0696	0,0158	0,9180	0,6981	0,0808	0,0102	0,8670
$\sigma_m^*$	0,8559	0,1139	0,0196	0,9320	1,2041	0,1139	0,0287	0,9240	1,0033	0,1616	0,0410	0,9260
$s_m$	1,8528	0,3669	0,2663	0,9590	2,2215	0,3669	0,5176	0,9510	1,4921	0,2742	0,1471	0,9560

Com o objetivo de verificar quantos termos autorregressivos devem ser incorporados no modelo para que a série fique melhor explicada, apresentamos um estudo de simulação que segue os seguintes passos:

Tabela 4.4: Estudo de simulação do modelo NG-PAR considerando as métricas VAM, ME, EQM e PC para  $m = 7, 8$  e  $9$ 

	m=7				m=8				m=9			
	ME	VAM	EQM	PC	ME	VAM	EQM	PC	ME	VAM	EQM	PC
	R=30				R=30				R=30			
$\phi_{1,m}$	0,7897	0,1904	0,0582	0,8130	0,7481	0,0918	0,0132	0,8560	0,5996	0,0412	0,0027	0,8880
$\sigma_m^*$	1,7298	0,4108	0,2757	0,8970	1,2536	0,2470	0,1048	0,9130	0,5736	0,0980	0,0167	0,9340
$s_m$	2,3596	0,9940	2,6699	0,9130	2,7531	1,1111	3,1707	0,9390	3,2872	1,3766	5,0040	0,9270
	R=50				R=50				R=50			
$\phi_{1,m}$	0,7966	0,1429	0,0327	0,8960	0,7504	0,0651	0,0070	0,9100	0,6013	0,0300	0,0014	0,9130
$\sigma_m^*$	1,8078	0,3295	0,1748	0,8900	1,2834	0,1969	0,0614	0,9190	0,5914	0,0775	0,0096	0,9250
$s_m$	2,2568	0,7700	1,9848	0,9460	2,5883	0,8681	2,3269	0,9530	3,2083	1,1600	3,9350	0,9460
	R=80				R=80				R=80			
$\phi_{1,m}$	0,8044	0,1157	0,0205	0,9180	0,7469	0,0525	0,0043	0,9210	0,5993	0,0236	0,0009	0,9190
$\sigma_m^*$	1,8055	0,2540	0,1008	0,9230	1,2885	0,1536	0,0386	0,9320	0,5963	0,0602	0,0058	0,9270
$s_m$	2,0107	0,4715	0,5296	0,9510	2,3900	0,5996	0,8972	0,9490	2,9468	0,8186	1,9080	0,9470
	R=100				R=100				R=100			
$\phi_{1,m}$	0,7952	0,1031	0,0163	0,9270	0,7502	0,0464	0,0034	0,9260	0,5983	0,0211	0,0007	0,9410
$\sigma_m^*$	1,8086	0,2207	0,0761	0,9340	1,3033	0,1363	0,0299	0,9320	0,5959	0,0545	0,0048	0,9230
$s_m$	1,9595	0,3994	0,3208	0,9600	2,3410	0,5026	0,6250	0,9670	2,7978	0,6795	1,1947	0,9450

Tabela 4.5: Estudo de simulação do modelo NG-PAR considerando as métricas VAM, ME, EQM e PC para  $m = 10, 11$  e  $12$ 

	m=10				m=11				m=12			
	ME	VAM	EQM	PC	ME	VAM	EQM	PC	ME	VAM	EQM	PC
	R=30				R=30				R=30			
$\phi_{1,m}$	0,4981	0,0457	0,0034	0,9060	0,2934	0,3088	0,1561	0,7850	0,5538	0,0618	0,0060	0,9060
$\sigma_m^*$	0,4731	0,0712	0,0086	0,9230	1,3618	0,3615	0,2033	0,8920	0,7077	0,1076	0,0201	0,9330
$s_m$	4,1956	1,8499	8,4437	0,9090	2,0733	0,8405	2,3592	0,9200	3,7939	1,5661	6,3293	0,9160
	R=50				R=50				R=50			
$\phi_{1,m}$	0,5017	0,0356	0,0021	0,8990	0,2933	0,2328	0,0885	0,8200	0,5503	0,0457	0,0034	0,9120
$\sigma_m^*$	0,4919	0,0528	0,0047	0,9280	1,3897	0,2973	0,1379	0,8850	0,7405	0,0815	0,0109	0,9050
$s_m$	4,3054	1,7175	8,1886	0,9480	1,9724	0,6814	2,0870	0,9300	4,0727	1,5970	7,1973	0,9570
	R=80				R=80				R=80			
$\phi_{1,m}$	0,4989	0,0271	0,0011	0,9300	0,2949	0,1792	0,0495	0,8890	0,5477	0,0354	0,0021	0,9170
$\sigma_m^*$	0,4973	0,0409	0,0027	0,9280	1,4180	0,2191	0,0743	0,9330	0,7538	0,0652	0,0068	0,9200
$s_m$	3,8946	1,1854	3,7496	0,9620	1,7760	0,3897	0,4248	0,9620	3,6854	1,0901	3,1536	0,9610
	R=100				R=100				R=100			
$\phi_{1,m}$	0,5000	0,0233	0,0009	0,9360	0,2904	0,1596	0,0406	0,9140	0,5485	0,0326	0,0017	0,9180
$\sigma_m^*$	0,4962	0,0377	0,0023	0,9330	1,3974	0,1941	0,0590	0,9370	0,7508	0,0578	0,0053	0,9260
$s_m$	3,7448	1,0231	2,8285	0,9510	1,7104	0,3226	0,1984	0,9520	3,5282	0,9229	2,4489	0,9640

1. gerar uma série com  $S = 12$ ,  $R = 30, 50, 80$  ou  $100$  e  $p_1 = \dots = p_{12} = 1, 2$  ou  $3$ ;
2. utilizando o processo de estimação descrito anteriormente obter o valor do máximo

- da função log-verossimilhança e a inversa da matriz de informação observada de Fisher para todas as possíveis combinações de  $m$  e  $p_m$ , sendo  $m = 1, \dots, 12$  e  $p_m = 1, 2$  e  $3$ ;
3. construir os intervalos de confiança dos parâmetros autorregressivos para todas as possíveis combinações de  $m$  e  $p_m$ ;
  4. verificar para cada combinação de valor de  $m$  e  $p_m$  se o parâmetro autorregressivo incluído a mais em relação ao cenário  $p_m = 1$  é significativo, para as combinações que não são significativos descartá-los;
  5. escolher a melhor ordem segundo os critérios AIC e BIC para todas as combinações que não foram descartadas no item anterior.
  6. repetir 1000 vezes os passos de 1 até 5.

Com os passos descritos temos que para cada combinação de  $m$ ,  $p_m$  e tamanho de série escolhidos no item 1, obtemos a quantidade de vezes que cada uma das três ordens adotadas no estudo foram indicadas como a melhor ordem para explicar as séries simuladas, isso para cada um dos critérios AIC e BIC. A quantidade de indicações, em porcentagem, de cada critério é apresentado nas Tabelas 4.7 a 4.12. Os valores dos parâmetros utilizados neste estudo estão apresentados na Tabela 4.6.

Tabela 4.6: Parâmetros adotados para o estudo de simulação do modelo NG-PAR para a análise dos critérios AIC e BIC

	m											
	1	2	3	4	5	6	7	8	9	10	11	12
$\phi_{1,m}$	0,60	0,70	0,45	-0,65	0,80	-0,75	0,39	-0,65	-0,45	-0,59	0,73	0,75
$\phi_{2,m}$	-0,45	-0,53	-0,82	0,47	-0,55	0,55	0,55	0,45	0,65	0,41	-0,59	-0,51
$\phi_{3,m}$	-0,50	0,43	0,63	0,56	0,42	0,35	-0,38	0,30	0,35	0,29	-0,39	0,41
$\sigma_m$	1,10	1,00	0,90	0,85	0,95	1,00	1,15	0,70	0,60	0,85	0,95	1,10
$s_m$	1,60	1,50	2,50	1,70	2,00	1,40	1,80	2,10	2,50	2,40	1,60	2,00

Observamos através dos resultados apresentados nas Tabelas 4.7 a 4.12, que de uma maneira geral a quantidade de acertos dos critérios cresce com o aumento do tamanho da série temporal. Quando a série foi gerada com um ou dois termos autorregressivos e com tamanho de 360 e 600, notamos que o acerto do critério AIC gira em torno de 70%,



Tabela 4.7: Porcentagem das indicativas da melhor ordem segundo os critérios AIC/BIC quando a geração é feita a partir do modelo NG-PAR para  $m = 1$  e 2

m=1					m=2				
Gerado	Estimado				Gerado	Estimado			
Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)	Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)
NG-PAR(1)	30	68,8/82,4	19,7/12,9	11,5/4,7	NG-PAR(1)	30	72,2/85,8	17,3/10,5	10,5/3,7
	50	82,7/94,3	11,1/4,0	6,2/1,7		50	80,5/95,3	12,1/3,5	7,4/1,2
	80	89,3/97,9	7,5/1,8	3,2/0,3		80	90,7/99,2	7,0/0,7	2,3/0,1
	100	92,1/98,4	6,0/1,3	1,9/0,3		100	90,2/98,7	7,5/1,3	2,3/0
NG-PAR(2)	30	12,0/20,3	69,6/69,6	18,4/10,1	NG-PAR(2)	30	11,9/23,8	67,6/66,4	20,5/9,8
	50	1,5/2,9	88,7/93,3	9,8/3,8		50	3,3/7,4	85,4/88,2	11,3/4,4
	80	1,0/1,1	92,9/96,9	6,1/2,0		80	1,2/2,1	88,7/96,2	10,1/1,7
	100	0/0	94,6/99,4	5,4/0,6		100	0,1/0,1	90,8/98,6	9,1/1,3
NG-PAR(3)	30	5,0/7,8	59,5/69,6	35,5/22,6	NG-PAR(3)	30	27,8/54,2	21,8/17,7	50,4/28,1
	50	0,9/1,1	55,7/71,7	43,4/27,2		50	15,5/42,2	9,9/11,4	74,6/46,4
	80	0,2/0,2	34,0/55,0	65,8/44,8		80	5,3/23,3	4,3/6,2	90,4/70,5
	100	0,1/0,1	18,9/42,1	81,0/57,8		100	2,5/14,8	4,0/3,9	93,5/81,3

Tabela 4.8: Porcentagem das indicativas da melhor ordem segundo os critérios AIC/BIC quando a geração é feita a partir do modelo NG-PAR para  $m = 3$  e 4

m=3					m=4				
Gerado	Estimado				Gerado	Estimado			
Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)	Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)
NG-PAR(1)	30	70,4/82,9	21,5/14,2	8,1/2,9	NG-PAR(1)	30	70,0/85,6	18,6/10,8	11,4/3,6
	50	87,6/95,7	8,2/3,3	4,2/1,0		50	82,9/93,7	11,3/5,5	5,8/0,8
	80	90,7/97,2	7,1/2,5	2,2/0,3		80	88,0/98,6	8,7/1,4	3,3/0
	100	92,5/98,5	6,2/1,3	1,3/0,2		100	92,1/99,2	5,8/0,7	2,1/0,1
NG-PAR(2)	30	33/38	84,2/87,8	12,5/8,4	NG-PAR(2)	30	8,0/12,8	74,3/77,3	17,7/9,9
	50	11/11	90,5/94,5	8,4/4,4		50	3,0/3,5	84,5/91,7	12,5/4,8
	80	4/4	93,5/97,5	6,1/2,1		80	1,4/1,5	89,4/96,5	9,2/2,0
	100	1/1	94,7/98,5	5,2/1,4		100	0,3/0,3	90,8/97,9	8,9/1,8
NG-PAR(3)	30	3,2/3,8	15,7/19,3	81,1/76,9	NG-PAR(3)	30	2,6/2,9	26,7/31,6	70,7/65,5
	50	1,1/1,1	7,5/7,8	91,4/91,1		50	0,5/0,5	9,0/10,7	90,5/88,8
	80	0/0	2,6/2,6	97,4/97,4		80	0/0	2,4/2,5	97,6/97,5
	100	0/0	1,5/1,5	98,5/98,5		100	0/0	1,5/1,5	98,5/98,5

Tabela 4.9: Porcentagem das indicativas da melhor ordem segundo os critérios AIC/BIC quando a geração é feita a partir do modelo NG-PAR para  $m = 5$  e 6

m=5					m=6				
Gerado	Estimado				Gerado	Estimado			
Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)	Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)
NG-PAR(1)	30	66,7/80,3	22,4/15,7	10,9/4,0	NG-PAR(1)	30	67,0/85,4	20,9/10,2	12,1/4,4
	50	85,7/95,4	10,1/3,7	4,2/0,9		50	80,7/94,9	11,4/3,6	7,9/1,5
	80	91,3/98,9	6,2/1,0	2,5/0,1		80	87,1/98,5	9,0/1,5	3,9/0
	100	92,3/99,1	5,9/0,8	1,8/0,1		100	90,1/99,2	7,7/0,8	2,2/0
NG-PAR(2)	30	15,6/26,7	68,4/64,9	16,0/8,4	NG-PAR(2)	30	30,9/56,1	50,3/36,6	18,8/7,3
	50	4,9/7,9	85,2/88,0	9,9/4,1		50	20,1/44,3	65,4/51,8	14,5/3,9
	80	1,1/2,1	91,6/96,8	7,3/1,1		80	14,4/30,4	75,5/67,6	10,1/2,0
	100	0,3/0,3	93,4/98,7	6,3/1,0		100	5,9/16,1	85,6/82,1	8,5/1,8
NG-PAR(3)	30	28,2/52,1	21,6/18,3	50,2/29,6	NG-PAR(3)	30	36,6/63,3	32,6/24,6	30,8/12,1
	50	13,2/37,6	13,4/13,6	73,4/48,8		50	24,3/51,6	29,8/26,7	45,9/21,7
	80	3,3/12,2	4,0/6,0	92,7/81,8		80	12,7/29,6	25,5/32,8	61,8/37,6
	100	0,9/5,0	1,6/3,1	97,5/91,9		100	7,3/19,8	22,0/31,5	70,7/48,7

enquanto do critério BIC fica em torno de 80%. Para estas ordens e com séries de tamanho igual a 1200, na maioria dos casos, o acerto dos critérios ultrapassa os 90% para ambos

Tabela 4.10: Porcentagem das indicativas da melhor ordem segundo os critérios AIC/BIC quando a geração é feita a partir do modelo NG-PAR para  $m = 7$  e  $8$ 

m=7					m=8				
Gerado	Estimado				Gerado	Estimado			
Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)	Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)
NG-PAR(1)	30	73,6/84,7	15,7/11,2	10,7/4,1	NG-PAR(1)	30	72,5/88,4	17,5/9,0	10,0/2,6
	50	84,6/95,3	8,8/3,0	6,6/1,7		50	84,4/93,5	10,5/5,0	5,1/1,5
	80	90,9/98,2	6,1/1,3	3,0/0,5		80	89,5/98,2	7,5/1,6	3,0/0,2
	100	91,1/98,8	7,0/1,2	1,9/0		100	89,5/98,4	8,3/1,5	2,2/0,1
NG-PAR(2)	30	12,2/14,2	72,6/76,2	15,2/9,6	NG-PAR(2)	30	15,5/16,5	71,8/75,8	12,7/7,7
	50	2,2/2,5	87,6/93,7	10,2/3,8		50	4,9/4,9	85,3/91,3	9,8/3,8
	80	0,8/0,8	92,5/97,7	6,7/1,5		80	0,7/0,7	91,5/97,0	7,8/2,3
	100	0,1/0,1	92,8/99,1	7,1/0,8		100	0,2/0,2	93,9/98,8	5,9/1,0
NG-PAR(3)	30	18,7/26,0	47,8/51,4	33,5/22,6	NG-PAR(3)	30	24,4/35,8	11,4/9,2	64,2/55,0
	50	5,5/10,7	46,9/58,1	47,6/31,2		50	8,3/12,2	5,4/3,3	86,3/84,5
	80	0,8/2,7	39,8/61,2	59,4/36,1		80	0,5/0,8	1,1/0,8	98,4/98,4
	100	0,4/2,0	27,1/52,5	72,5/45,5		100	0,1/0,1	0,5/0,5	99,4/99,4

Tabela 4.11: Porcentagem das indicativas da melhor ordem segundo os critérios AIC/BIC quando a geração é feita a partir do modelo NG-PAR para  $m = 9$  e  $10$ 

m=9					m=10				
Gerado	Estimado				Gerado	Estimado			
Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)	Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)
NG-PAR(1)	30	75,2/87,7	17,2/10,0	7,6/2,3	NG-PAR(1)	30	73,1/85,4	18,0/10,9	8,9/3,7
	50	84,8/94,8	10,5/3,9	4,7/1,3		50	85,3/94,5	9,8/4,8	4,9/0,7
	80	91,2/98,6	7,6/1,2	1,2/0,2		80	88,5/97,9	8,2/2,0	3,3/0,1
	100	92,8/98,2	6,3/1,6	0,9/0,2		100	91,4/99,2	6,9/0,7	1,7/0,1
NG-PAR(2)	30	17,7/18,2	71,8/75,8	10,5/6,0	NG-PAR(2)	30	44,1/59,5	47,6/36,8	8,3/3,7
	50	5,1/5,1	85,6/91,6	9,3/3,3		50	24,7/42,4	69,2/54,5	6,1/3,1
	80	0,7/0,7	91,6/97,9	7,7/1,4		80	8,2/21,6	87,1/77,8	4,7/0,6
	100	0,2/0,2	91,1/98,0	8,7/1,8		100	1,7/9,0	93,1/90,3	5,2/0,7
NG-PAR(3)	30	18,5/18,7	17,1/17,3	64,4/64,0	NG-PAR(3)	30	36,9/54,2	39,3/33,0	23,8/12,8
	50	8,7/8,7	8,5/8,5	82,8/82,8		50	15,4/32,0	39,5/41,9	45,1/26,1
	80	1,0/1,0	2,8/2,8	96,2/96,2		80	2,8/11,3	26,3/41,8	70,9/46,9
	100	0,2/0,2	0,5/0,5	99,3/99,3		100	1,1/5,9	15,9/34,8	83,0/59,3

Tabela 4.12: Porcentagem das indicativas da melhor ordem segundo os critérios AIC/BIC quando a geração é feita a partir do modelo NG-PAR para  $m = 11$  e  $12$ 

m=11					m=12				
Gerado	Estimado				Gerado	Estimado			
Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)	Modelo	R	NG-PAR(1)	NG-PAR(2)	NG-PAR(3)
NG-PAR(1)	30	71,4/86,2	17,1/10,0	11,5/3,8	NG-PAR(1)	30	73,5/87,7	16,3/9,3	10,2/3,0
	50	84,3/95,2	8,5/3,6	7,2/1,2		50	84,5/93,9	10,2/5,0	5,3/1,1
	80	87,9/98,3	9,2/1,7	2,9/0		80	91,1/99,0	6,6/0,9	2,3/0,1
	100	89,0/98,5	7,7/1,5	3,3/0		100	93,0/98,8	5,7/1,1	1,3/0,1
NG-PAR(2)	30	42,1/61,6	45,3/33,2	12,6/5,2	NG-PAR(2)	30	46,1/67,2	44,9/28,8	9,0/4,0
	50	18,9/34,2	71,2/63,8	9,9/2,0		50	24,0/42,4	68,3/55,8	7,7/1,8
	80	6,2/14,2	85,1/83,7	8,7/2,1		80	8,9/25,1	83,0/73,0	8,1/1,9
	100	1,9/6,3	89,4/92,2	8,7/1,5		100	2,8/12,7	92,0/86,5	5,2/0,8
NG-PAR(3)	30	32,8/50,0	34,5/31,9	32,7/18,1	NG-PAR(3)	30	19,0/28,3	55,0/55,7	26,0/16,0
	50	11,8/22,0	35,6/44,6	52,6/33,4		50	7,0/13,0	45,7/56,4	47,3/30,6
	80	1,3/3,9	24,3/43,3	74,4/52,8		80	0,6/3,1	23,6/36,7	75,8/60,2
	100	0,5/1,1	16,8/38,8	82,7/60,1		100	0,3/1,1	11,6/22,3	88,1/76,6

os critérios, constatamos ainda que o critério BIC apresenta maior quantidade de acertos em comparação ao outro critério.

Verificamos também que o critério AIC apresenta melhor desempenho do que o critério BIC quando geramos séries com três termos autorregressivos, observamos ainda dificuldades dos critérios em escolher o modelo com três termos quando a série é de tamanho igual a 360 ou 600. Esta dificuldade é amenizada quando o tamanho da série passa a ser de 960 e 1200.

## 4.5 Aplicação à Série de Furnas

Adotamos a série de vazões médias mensais afluentes da usina hidrelétrica de Furnas, apresentada na Seção 2.3.1, e retiramos a sazonalidade utilizando a equação (4.1) e adotando  $\hat{\mu}_m$  e  $\hat{\sigma}_m$  dados pelas equações (1.5) e (1.6). Investigamos a existência de correlação periódica por meio de gráficos de dispersão de diferentes períodos, pois segundo Hurd e Gerr (1991) a correlação periódica pode ser investigada desta maneira. Na Figura 4.1 apresentamos os gráficos de dispersão para alguns meses.

Observamos através da Figura 4.1 que as correlações entre diferentes períodos são diferentes, com isso existe indícios da existência de correlação periódica. Para uma melhor análise utilizamos as funções de autocorrelação periódica e autocorrelação parcial periódica, que são apresentadas na Figura 4.2.

Através da Figura 4.2 observamos que a série em estudo apresenta função de autocorrelação que não é igual em todos os meses, indicando assim que o modelo NG-PAR é indicado para modelar a série em estudo. Por meio da função de autocorrelação parcial periódica podemos ter ideia das ordens do modelo periódico autorregressivo que são (1, 1, 1, 2, 3, 1, 2, 1, 3, 3, 1, 2) para os meses de janeiro a dezembro, respectivamente.

A fim de escolher as melhores ordens que representam a série de Furnas, ajustamos para cada mês,  $m = 1, \dots, 12$ , diferentes ordens do modelo NG-PAR. Utilizando os critérios de seleção AIC e BIC escolhemos a melhor ordem para cada  $m$ . Os resultados dos critérios para as diferentes ordens e para todos os meses estão apresentados na Tabela 4.13.

Note que para os meses de julho e outubro aparecem \*\*\* na ordem 2, pois essa foi a ordem escolhida mesmo que ambos os critérios indicassem a ordem superior. A decisão

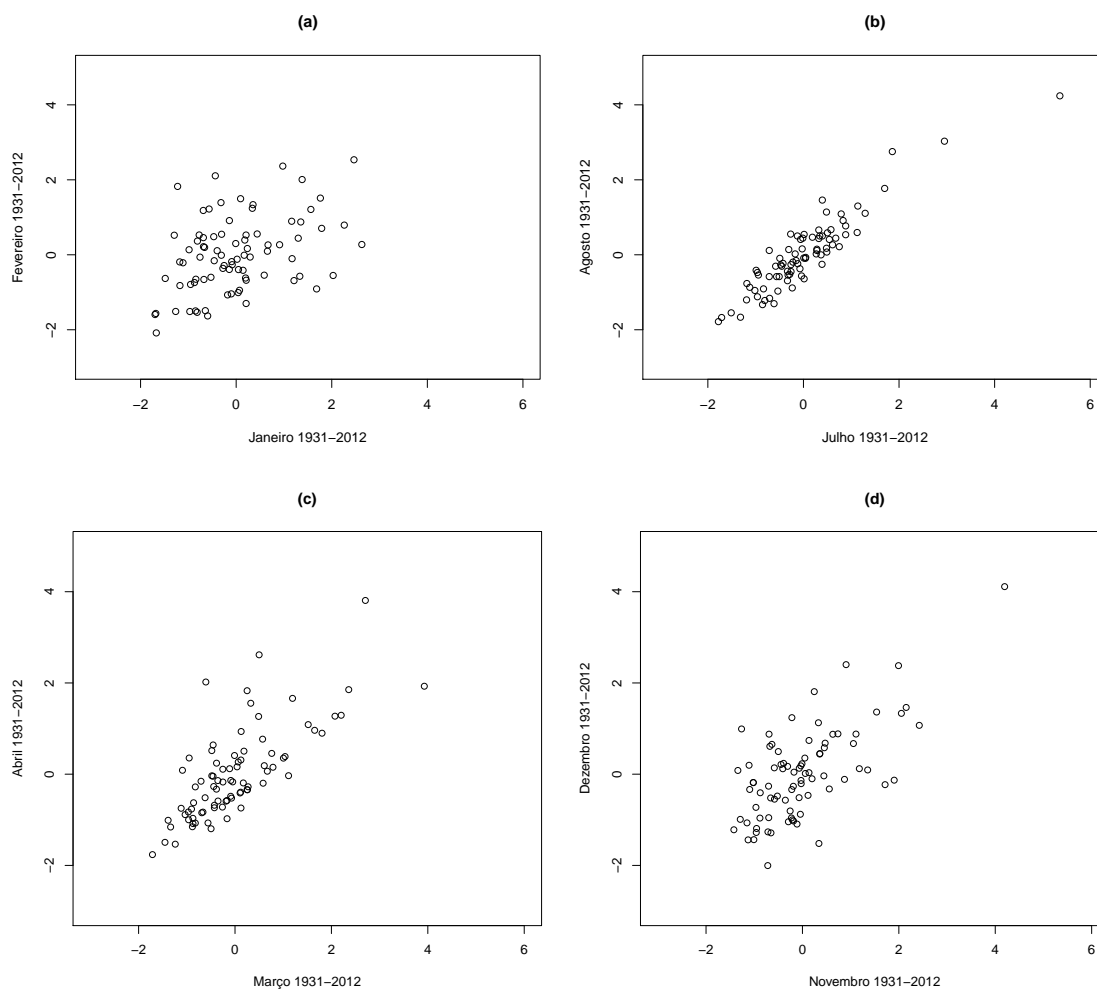


Figura 4.1: Gráfico de dispersão entre os meses de janeiro e fevereiro em (a), entre julho e agosto em (b), entre março e abril em (c) e entre novembro e dezembro em (d) da série de Furnas

de escolher essas ordens é devido ao fato que o ajuste com a ordem superior apresentou parâmetro(s) não significativo(s). Observe também que para os meses de fevereiro, março, abril e novembro aparecem \* nas ordens 1, 1, 2 e 1, respectivamente, isso significa que escolhemos essas ordens, pois cada critério indicou uma ordem e decidimos pela ordem que apresenta menor quantidade de parâmetros. Nos demais meses as ordens escolhidas são as indicadas através dos critérios.

Comparando as ordens resultantes da aplicação dos critérios AIC e BIC e das obtidas por meio da análise da função de autocorrelação parcial periódica, notamos que houve divergência apenas nos meses de junho, setembro e outubro, sendo que a análise gráfica forneceu ordens 1, 3 e 3 enquanto através dos critérios obtemos as ordens 3, 1 e 2,

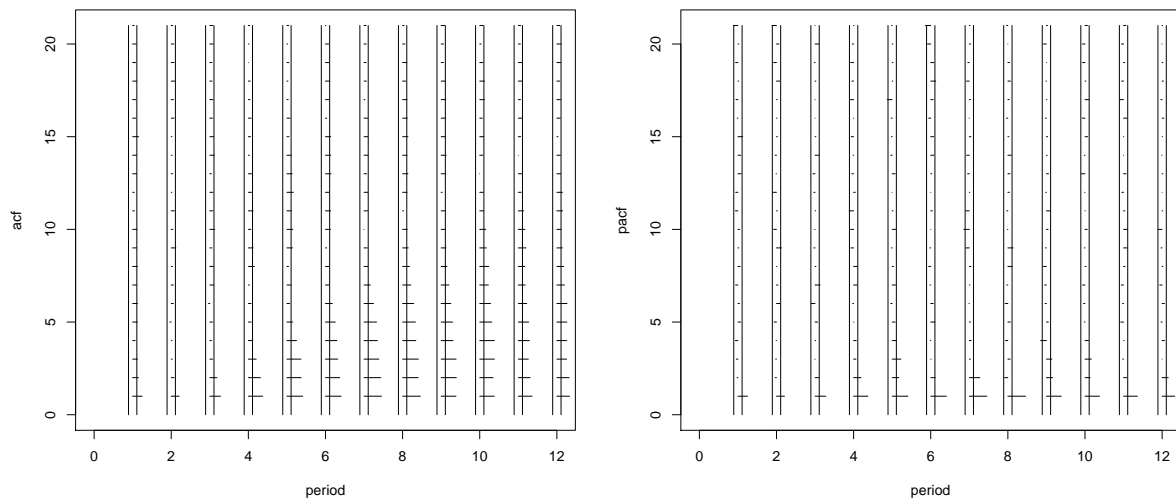


Figura 4.2: Funções de autocorrelação periódica (esquerda) e autocorrelação periódica parcial (direita) da série de Furnas

Tabela 4.13: Seleção da melhor ordem do modelo NG-PAR utilizando os critérios AIC e BIC para explicar a série de Furnas

m=1			m=2			m=3		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>209,18</b>	<b>216,36</b>	1*	218,69	<b>225,87</b>	1*	197,18	<b>204,37</b>
2	211,16	220,73	2	<b>211,75</b>	221,33	2	<b>195,51</b>	205,09
3	212,32	224,30	3	213,28	225,25	3	196,03	208,00
m=4			m=5			m=6		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	165,38	172,57	1	141,97	149,15	1	78,81	85,99
2*	152,84	<b>162,41</b>	2	136,43	146,01	2	78,21	87,78
3	<b>152,76</b>	164,73	3	<b>116,92</b>	<b>128,89</b>	3	<b>70,34</b>	<b>82,31</b>
m=7			m=8			m=9		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	83,49	90,67	1	<b>80,97</b>	<b>88,15</b>	1	<b>142,29</b>	<b>149,47</b>
2***	57,40	66,98	2	81,19	90,76	2	143,58	153,16
3	<b>54,11</b>	<b>66,09</b>	3	81,21	93,18	3	144,70	156,67
m=10			m=11			m=12		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	166,54	173,73	1*	177,41	<b>184,59</b>	1	189,32	196,50
2***	162,06	171,64	2	<b>176,85</b>	186,43	2	<b>182,81</b>	<b>192,38</b>
3	<b>153,17</b>	<b>165,14</b>	3	177,47	189,44	3	184,43	196,40

respectivamente para os meses de junho, setembro e outubro.

Na Tabela 4.14 apresentamos as estimativas de máxima verossimilhança e os intervalos de confiança dos parâmetros das ordens escolhidas. Observamos que todos os

parâmetros são significativos, mas note também que para os meses de janeiro, fevereiro, agosto, outubro, novembro e dezembro o intervalo de confiança para o parâmetro  $s$  contém o valor 2, implicando assim que a distribuição normal pode ser utilizada nesses meses. Para os demais meses, os intervalos de confiança para  $s$  não contém o valor 2, indicando assim que a distribuição normal generalizada é a mais indicada para o ajuste, havendo assim um empate entre as distribuições normal e normal generalizada, pois cada uma delas é indicada no ajuste de seis meses.

Tabela 4.14: Estimativas pontual e intervalar dos parâmetros do modelo NG-PAR que melhor se ajusta à série de Furnas

	EMV	IC	EMV	IC	EMV	IC
	Janeiro		Fevereiro		Março	
$\phi_1$	0,5095	(0,3341; 0,6849)	0,4305	(0,2339; 0,6271)	0,5646	(0,4469; 0,6823)
$\sigma$	1,3424	(1,0319; 1,6529)	1,3227	(0,9324; 1,7131)	0,7827	(0,4607; 1,1047)
$s$	2,4876	(1,2710; 3,7041)	2,1667	(0,9678; 3,3655)	1,2462	(0,8090; 1,6835)
	Abril		Maio		Junho	
$\phi_1$	0,5576	(0,3640; 0,7512)	0,4595	(0,2521; 0,6668)	0,4950	(0,4845; 0,5054)
$\phi_2$	0,3080	(0,1715; 0,4444)	0,2531	(0,0912; 0,4150)	0,1536	(0,1399; 0,1672)
$\phi_3$			0,2865	(0,1374; 0,4356)	0,1076	(0,0931; 0,1222)
$\sigma$	0,5603	(0,2930; 0,8276)	0,5068	(0,2852; 0,7284)	0,1386	(0,0220; 0,2553)
$s$	1,1912	(0,7230; 1,6594)	1,3613	(0,7684; 1,9542)	0,6986	(0,4678; 0,9294)
	Julho		Agosto		Setembro	
$\phi_1$	0,5437	(0,4259; 0,6616)	0,9215	(0,8348; 1,0083)	0,5569	(0,5183; 0,5954)
$\phi_2$	0,4188	(0,2742; 0,5634)				
$\sigma$	0,3003	(0,1547; 0,4459)	0,5298	(0,3752; 0,6843)	0,3998	(0,1875; 0,6121)
$s$	1,1553	(0,7138; 1,5968)	1,9066	(1,0644; 2,7487)	0,9568	(0,6527; 1,2609)
	Outubro		Novembro		Dezembro	
$\phi_1$	0,5482	(0,3526; 0,7439)	0,7122	(0,5735; 0,8509)	0,3929	(0,1606; 0,6252)
$\phi_2$	0,2709	(0,0692; 0,4725)			0,3459	(0,1191; 0,5726)
$\sigma$	0,9453	(0,7249; 1,1656)	0,9232	(0,6431; 1,2032)	1,0715	(0,7680; 1,3750)
$s$	2,2973	(1,3026; 3,2921)	1,7805	(1,0386; 2,5224)	2,2827	(0,9823; 3,5831)

A fim de comparação ajustamos o modelo periódico autorregressivo para todos os meses utilizando a distribuição normal. Apresentamos os resultados dos critérios AIC e BIC na Tabela 4.15, já as estimativas de máxima verossimilhança dos parâmetros para as ordens escolhidas são apresentadas na Tabela 4.16.

Tabela 4.15: Seleção da melhor ordem do modelo N-PAR utilizando os critérios AIC e BIC para explicar a série de Furnas

m=1			m=2			m=3		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>207,93</b>	<b>212,72</b>	1***	216,77	221,56	1***	203,02	207,81
2	209,88	217,06	2	<b>209,94</b>	<b>217,12</b>	2	<b>201,57</b>	<b>208,75</b>
3	211,40	220,98	3	211,47	221,04	3	202,00	211,58
m=4			m=5			m=6		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	169,62	174,41	1	144,97	149,76	1	145,64	150,43
2	<b>158,58</b>	<b>165,77</b>	2	136,13	143,31	2*	142,31	<b>149,49</b>
3	158,67	168,24	3	<b>117,77</b>	<b>127,35</b>	3	<b>140,07</b>	149,65
m=7			m=8			m=9		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	91,45	96,24	1	<b>79,01</b>	<b>83,80</b>	1***	161,11	165,90
2*	65,35	<b>72,54</b>	2	79,20	86,38	2	153,69	160,87
3	<b>63,31</b>	72,89	3	79,47	89,05	3	<b>148,22</b>	<b>157,80</b>
m=10			m=11			m=12		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	164,56	169,35	1*	175,71	<b>180,50</b>	1	187,36	192,15
2***	160,45	167,64	2	<b>175,51</b>	182,69	2	<b>181,01</b>	<b>188,20</b>
3	<b>151,18</b>	<b>160,76</b>	3	176,39	185,96	3	182,66	192,24

A utilização da \* e \*\*\* na Tabela 4.15 tem o mesmo significado empregado na Tabela 4.13. Observamos na Tabela 4.15 que a escolha das ordens usando os critérios AIC e BIC quando adotamos a distribuição normal só diferem das escolhas das ordens quando utilizamos a distribuição normal generalizada no mês de junho.

Analisando as estimativas de máxima verossimilhança dos parâmetros  $\phi$ 's, dos meses com as mesmas ordens quando utilizamos as distribuições normal e normal generalizada, observamos que exceto no mês de setembro as estimativas são próximas, pois a estimativa do  $\phi_1$  quando usamos a distribuição normal é de 0,7670 e quando adotamos a distribuição normal generalizada esse valor é de 0,5569. Acreditamos que o motivo da diferença desses valores é o comportamento das observações do mês de setembro que apresentam caudas que diferem em muito das caudas da distribuição normal.

Para a escolha entre as distribuições normal e normal generalizada utilizamos os critérios AIC e BIC, que por sua vez é a soma dos valores para cada mês. O resultado é apresentado na Tabela 4.17. Notamos que o modelo NG-PAR se ajusta melhor à série de vazões médias mensais da usina hidrelétrica de Furnas em comparação ao modelo PAR

Tabela 4.16: Estimativas pontual e intervalar dos parâmetros do modelo N-PAR que melhor se ajusta à série de Furnas

	EMV	IC	EMV	IC	EMV	IC
	Janeiro		Fevereiro		Março	
$\phi_1$	0,4969	(0,3056; 0,6883)	0,4334	(0,2357; 0,6312)	0,5540	(0,3735; 0,7346)
$\sigma$	0,8656	(0,7314; 0,9997)	0,8998	(0,7613; 1,0384)	0,8266	(0,6993; 0,9539)
	Abril		Maio		Junho	
$\phi_1$	0,5643	(0,4007; 0,7278)	0,4358	(0,2683; 0,6033)	0,6061	(0,3945; 0,8177)
$\phi_2$	0,3128	(0,1497; 0,4758)	0,2417	(0,0846; 0,3989)	0,2544	(0,0420; 0,4668)
$\phi_3$			0,3067	(0,1694; 0,4441)		
$\sigma$	0,6206	(0,5250; 0,7162)	0,4765	(0,4031; 0,5499)	0,5613	(0,4749; 0,6477)
	Julho		Agosto		Setembro	
$\phi_1$	0,5959	(0,4649; 0,72694)	0,9204	(0,8367; 1,0041)	0,7670	(0,6277; 0,9064)
$\phi_2$	0,3871	(0,2562; 0,5179)				
$\sigma$	0,3490	(0,2953; 0,4028)	0,3845	(0,3253; 0,4437)	0,6382	(0,5399; 0,7365)
	Outubro		Novembro		Dezembro	
$\phi_1$	0,5384	(0,3242; 0,7527)	0,7139	(0,5605; 0,8673)	0,4099	(0,1877; 0,6322)
$\phi_2$	0,2748	(0,0609; 0,4888)			0,3371	(0,1142; 0,5600)
$\sigma$	0,6278	(0,5311; 0,7245)	0,6984	(0,5908; 0,8059)	0,7128	(0,6030; 0,8225)

com distribuição normal.

Tabela 4.17: Resultado dos critérios AIC e BIC dos ajustes dos modelo NG-PAR e N-PAR à série de Furnas

	AIC	BIC
N-PAR	1869,02	1943,28
NG-PAR	<b>1768,09</b>	<b>1873,42</b>

Analisando os critérios de seleção em cada mês, observamos o fato que cada distribuição é indicada para o ajuste de seis meses. Consideramos que o melhor ajuste é aquele que mescla as duas distribuições, ou seja, é o ajuste que leva em conta a distribuição normal generalizada nos meses de março, abril, maio, junho, julho e setembro, e a distribuição normal nos demais meses. Os valores dos critérios AIC e BIC do modelo final é 1757,85 e 1848,85, respectivamente.

Para examinar a qualidade do ajuste, Mondal e Wasimi (2006) sugerem as análises das funções de autocorrelação periódica e autocorrelação parcial periódica dos resíduos,



e agrupar os resíduos e explorar as funções de autocorrelação e autocorrelação parcial.

A partir do modelo final e das estimativas de máxima verossimilhança obtidas calculamos os resíduos. Para examinar a qualidade do ajuste apresentamos na Figura 4.3 as funções de autocorrelação periódica e autocorrelação parcial periódica dos resíduos. Já na Figura 4.4 mostramos as funções de autocorrelação e autocorrelação parcial dos resíduos agrupados.

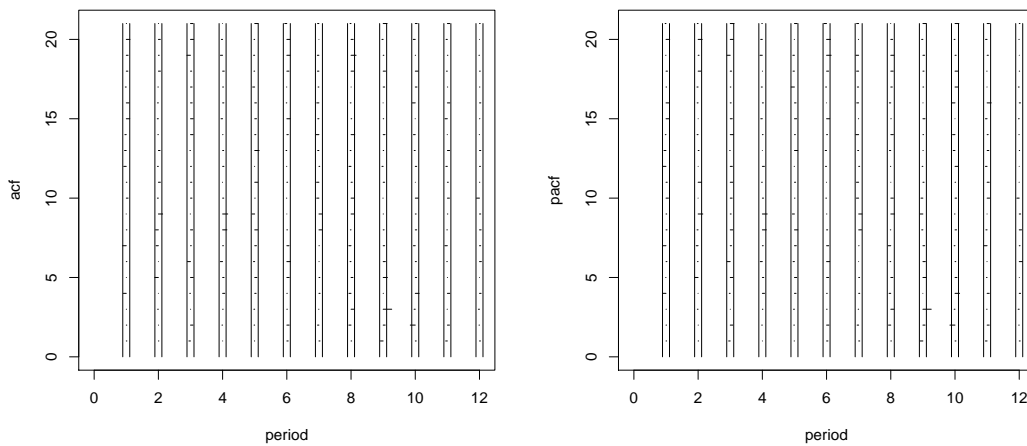


Figura 4.3: Funções de autocorrelação periódica (esquerda) e autocorrelação periódica parcial (direita) dos resíduos do ajuste do modelo final à série de Furnas

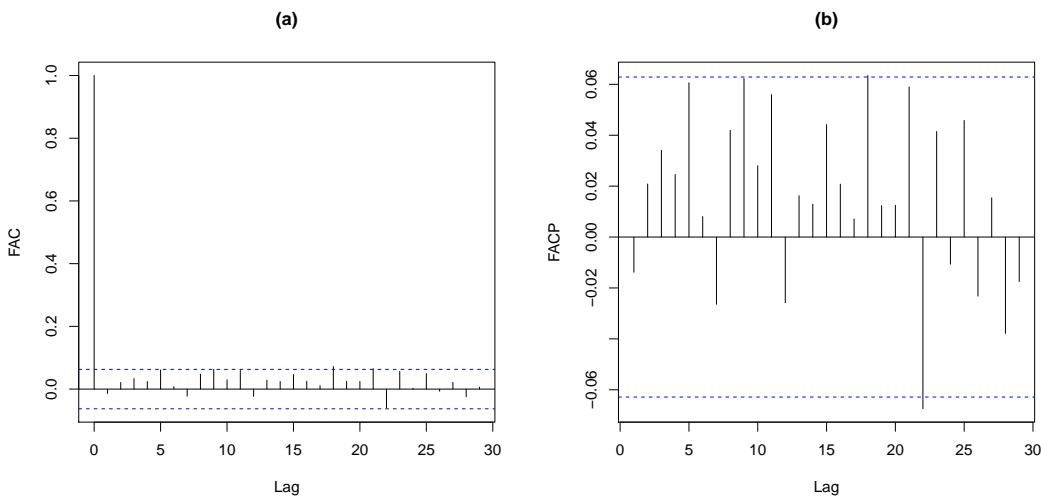


Figura 4.4: Funções de autocorrelação em (a) e autocorrelação parcial dos resíduos em (b) dos resíduos do ajuste do modelo final à série de Furnas

Observamos através das Figuras 4.3 e 4.4 que o ajuste eliminou a correlação

existente na série, indicando assim que houve um bom ajuste do modelo aos dados.

As previsões foram feitas considerando os dois cenários apresentados no Capítulo 2, lembrando que as previsões do segundo cenário utilizam mais informações que as previsões do primeiro cenário. As previsões são apresentadas na Figura 4.5.

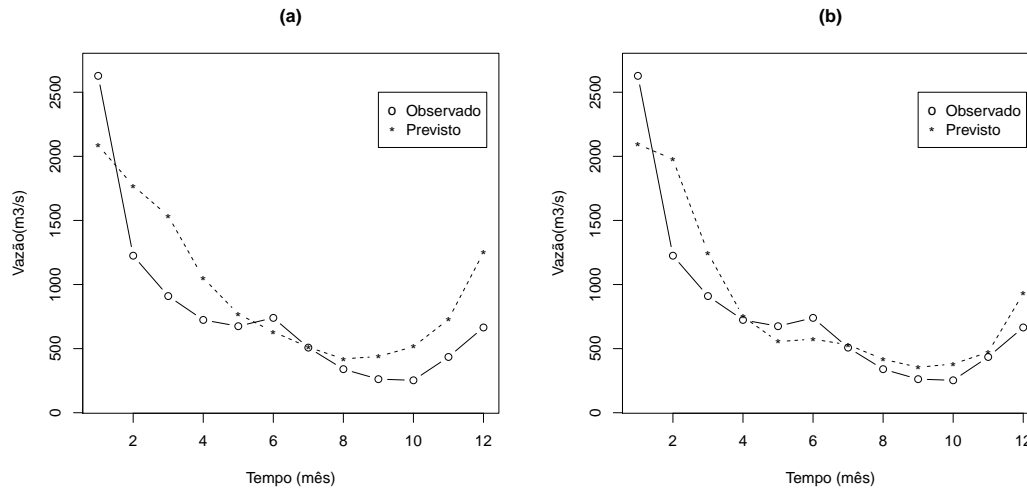


Figura 4.5: Previsões para a série de Furnas adotando o modelo PAR com distribuição condicional normal e normal generalizada e considerando o 1º cenário em (a) e o 2º cenário em (b)

Comparando as previsões apresentadas nas Figuras 2.7 e 4.5, notamos pequenas diferenças em ambos os cenários, para uma comparação mais justa comparamos as previsões dos modelos a partir dos resultados das métricas adotadas neste trabalho.

Através da Tabela 4.18 apresentamos os resultados dos critérios MAPE, EQM e EAM que medem a eficiência das previsões para o ano de 2012 da série de vazão de Furnas quando utilizamos o modelo NG-PAR. Comparando com os resultados obtidos a partir do modelo NG-ARMA, apresentados na Tabela 2.7, notamos que em ambos os cenários as previsões utilizando o modelo NG-PAR alcançaram melhores resultados nas três métricas utilizadas.

Tabela 4.18: Avaliação das previsões para a série de Furnas feitas a partir do modelo PAR com distribuição condicional normal e normal generalizada segundo os critérios MAPE, EQM e EAM

Cenário	MAPE(%)	EQM	EAM
1º	46,5	137.053,7	303,7
2º	26,8	92.918,9	212,6

## 4.6 Conclusão

Com o interesse em uma maior robustez, apresentamos neste capítulo uma modificação dos modelos PAR substituindo a distribuição normal, usualmente utilizada, pela normal generalizada, aderindo assim ao modelo uma maior flexibilidade no ajuste de séries que apresentam simetria, mas com caudas que podem ser mais leves ou mais pesadas do que as da distribuição normal.

Mostramos também neste capítulo a função de verossimilhança condicional. Desenvolvemos dois estudos de simulações para verificar as propriedades dos estimadores de máxima verossimilhança e o comportamento dos critérios de seleção AIC e BIC. A partir do modelo NG-PAR mostramos um método para o cálculo de previsões.

Exemplificamos uma aplicação do modelo proposto ajustando-o à série de vazões médias mensais afluentes da barragem na usina hidrelétrica de Furnas. Comparando os ajustes dos modelos NG-ARMA e NG-PAR à série de Furnas, notamos que as previsões para o modelo NG-PAR apresentaram melhores resultados.

# Capítulo 5

## Modelo Assimétrico Normal

## Generalizada Autorregressivo Periódico

O modelo desenvolvido no Capítulo 4 flexibiliza o tradicional modelo autorregressivo periódico com distribuição normal no sentido de melhor explicar séries que apresentam periodicidade e distribuição simétrica com curtose diferente da distribuição normal. A fim de estender o modelo autorregressivo periódico para que possa ser utilizado para séries que possuam distribuição assimétrica, neste trabalho adotamos duas alternativas, a primeira é a utilização da distribuição log-normal generalizada como a distribuição marginal, a segunda é fazer o uso da transformação de Box-Cox para atenuar a assimetria da série temporal.

Neste capítulo apresentamos os modelos log-normal generalizada autorregressivo periódico (LNG-PAR) e o Box-Cox normal generalizada autorregressivo periódico (BC-NG-PAR) como sendo alternativas para a modelagem de séries temporais que apresentam comportamento periódico e distribuição assimétrica. Na Seção 5.1 apresentamos o modelo LNG-PAR, juntamente com um procedimento de previsão através do modelo proposto e uma aplicação com a série temporal de vazões da usina de Furnas. O modelo BC-NG-PAR é exposto na Seção 5.2 em conjunto com a função verossimilhança, um método para o cálculo de previsões e com duas aplicações, adotando as séries de vazões das usinas de Furnas e Sobradinho. Na Seção 5.3 apresentamos alguns comentários.

## 5.1 Modelo Log-Normal Generalizada Autorregressivo Periódico

Consideremos  $\{X_{t(r,m)}; r = 1, \dots, R; m = 1, \dots, S\}$  uma série temporal periódica, assumindo que a distribuição da série temporal, para todo instante de tempo, condicional em todo o passado seja a distribuição log-normal generalizada com os parâmetros  $\mu_m^*$ ,  $\sigma_m^*$  e  $s_m$ , sendo que  $\mu_m^*$  é dado pela equação (4.3).

Definimos o modelo LNG-PAR através da equação (4.3) e da utilização da distribuição log-normal generalizada como distribuição condicional. O modelo proposto é apresentado como uma alternativa para modelar séries temporais que apresentam periodicidade, assimetria e com valores nos reais positivos.

Eliminamos a sazonalidade da série observada adotando a equação

$$Z_{t(r,m)} = \frac{\log(X_{t(r,m)}) - \mu_m}{\sigma_m}, \quad (5.1)$$

sendo que  $\mu_m = E\{\log(X_{t(r,m)})\}$ ,  $\sigma_m = \sqrt{Var\{\log(X_{t(r,m)})\}}$  e  $\{Z_{t(r,m)}; r = 1, \dots, R; m = 1, \dots, S\}$  é uma série temporal que exhibe comportamento periódico na função de correlação. Comparando as equações (3.3) e (5.1) observamos que a diferença entre elas está no índice de tempo da série  $Z_t$ , ou seja, no Capítulo 3 não fizemos nenhuma suposição a cerca da série livre de sazonalidade, sendo que neste capítulo impomos que a série livre de sazonalidade apresenta função de correlação periódica.

Observemos que na equação (5.1) aplicamos a função logaritmo na série observada, o que resulta que a série livre de sazonalidade apresenta distribuição marginal normal generalizada, pois o modelo LNG-PAR é definido utilizando a distribuição condicional log-normal generalizada. Portanto, para utilizarmos o modelo LNG-PAR basta eliminar a sazonalidade da série original usando a equação (5.1) e adotar tudo o que foi desenvolvido para o modelo NG-PAR, apresentado no Capítulo 4.

### 5.1.1 Previsão

Nesta Seção o nosso interesse é desenvolver um procedimento para prever valores futuros da série  $X_t$ , ou seja, estamos empenhados em prever um valor da série original no

instante  $t+h$  com  $h \geq 1$ , supondo que temos as observações até o instante  $t$ . Para prever valores para a série original precisamos inicialmente profetizar valores para a série livre de sazonalidade. Desde que a equação (5.1) seja usada para eliminar a sazonalidade da série original, como apresentado na Seção 5.1, utilizamos o modelo NG-PAR para inferir sobre o modelo LNG-PAR. Logo, a partir do modelo NG-PAR conseguimos prever valores livres de sazonalidade a partir das equações (4.7) e (4.9).

A partir dos valores preditos do modelo NG-PAR conseguimos prever os valores para a série original. Optamos por prever através da mediana para que não precisemos fazer aproximações, para isso definimos

$$P(Z_{t(r,m)+h} \leq z_{t(r,m)+h}^{(M)} | \mathfrak{F}_t) = 0,5,$$

assim  $z_{t(r,m)+h}^{(M)}$  é o valor da mediana da distribuição  $Z_{t(r,m)+h} | \mathfrak{F}_t$ , mas podemos reescrever a série  $Z_{t(r,m)}$  como função da série  $X_{t(r,m)}$ , disto temos que

$$P\left(\frac{\log(X_{t(r_0,m_0)}) - \hat{\mu}_{m_0}}{\hat{\sigma}_{m_0}} \leq z_{t+h}^{(M)} | \mathfrak{F}_t\right) = 0,5,$$

em que  $t(r, m) + h = t(r_0, m_0)$ . Definindo  $\mu_t^*(h)$  como a previsão para  $X_{t(r,m)+h}$  temos que

$$\mu_t^*(h) = \exp[z_{t+h}^{(M)} \hat{\sigma}_{m_0} + \hat{\mu}_{m_0}],$$

por construção  $Z_t | \mathfrak{F}_{t-1}$  segue uma distribuição normal generalizada então a mediana é igual à média. Mas predizemos a série livre de sazonalidade a partir das equações (4.7) e (4.9) utilizando a média condicional, logo a mediana é igual à média predita através do modelo NG-PAR. Definindo  $\hat{\mu}_t^*(h)$  como a previsão para  $X_{t+h}$ , então

$$\mu_t^*(h) = \exp[\hat{\mu}_{t+h} \hat{\sigma}_{m_0} + \hat{\mu}_{m_0}], \quad (5.2)$$

sendo que  $\hat{\mu}_{t+h}$  é a predição através do modelo NG-PAR, a equação é válida para qualquer  $h \geq 1$ .

### 5.1.2 Aplicação à Série de Furnas

Adotamos a série de vazões médias mensais afluentes da usina hidrelétrica de Furnas apresentada na Seção 2.3.1 e eliminamos a sazonalidade utilizando a equação (5.1).

Utilizamos as funções de autocorrelação periódica e autocorrelação parcial periódica para investigar a existência de correlação periódica, as funções são apresentada na Figura 5.2.

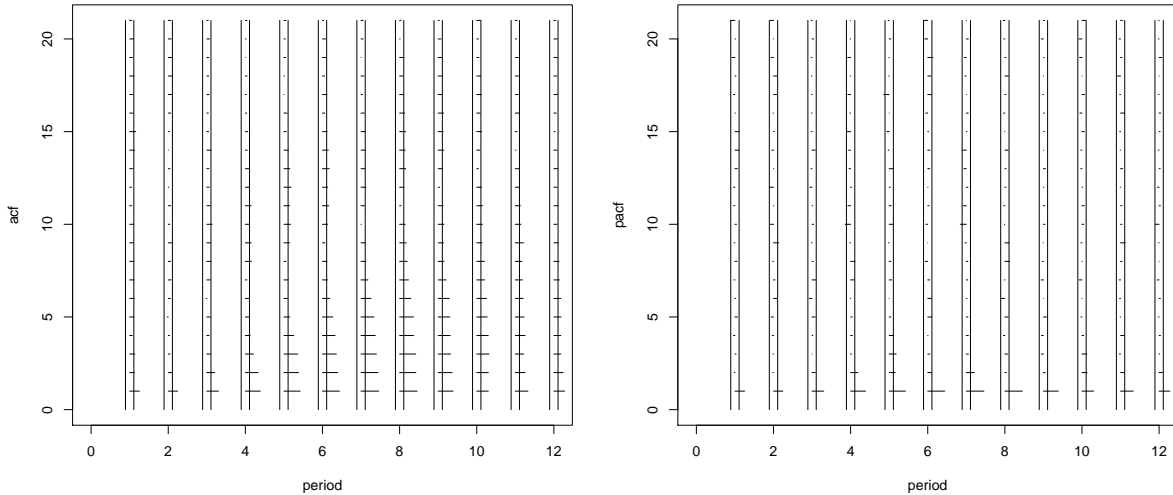


Figura 5.1: Funções de autocorrelação periódica (esquerda) e autocorrelação periódica parcial (direita)

Observamos por meio da Figura 5.2 que a série em estudo apresenta correlação e correlação parcial diferente para cada mês. Como os valores observados da série são positivos então o modelo LNG-PAR é indicado para modelar a série em estudo. Por meio da função de autocorrelação parcial periódica podemos ter ideia das ordens do modelo periódico autorregressivo, que são  $(1, 1, 1, 2, 3, 1, 2, 1, 1, 3, 1, 1)$  para os meses de janeiro a dezembro, respectivamente.

Escolhemos as melhores ordens que representam a série de Furnas ajustando para cada mês diferentes ordens do modelo LNG-PAR. Utilizando a combinação entre verificação de significância dos parâmetros autorregressivos e os critérios de seleção AIC e BIC, escolhemos a melhor ordem para cada mês. Os resultados dos critérios para as diferentes ordens e para todos os  $m$  estão apresentados na Tabela 5.1.

O significado dos símbolos \* e \*\*\* na Tabela 5.1 é o mesmo empregado na Tabela 4.13. Comparando as ordens escolhidas e das indicadas por meio da análise da função de autocorrelação parcial periódica observamos que houve divergência apenas no mês de outubro, sendo que a análise gráfica forneceu a ordem 3 enquanto através dos critérios

Tabela 5.1: Seleção dos modelos utilizando os critérios AIC e BIC quando adotamos a distribuição log-normal generalizada

Janeiro (m=1)			Fevereiro (m=2)			Março (m=3)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>207,58</b>	<b>214,76</b>	1***	212,46	219,65	1*	197,04	<b>204,22</b>
2	209,25	218,83	2	<b>208,93</b>	<b>218,51</b>	2	<b>195,63</b>	205,21
3	210,22	222,20	3	210,74	222,71	3	196,03	208,00
Abril (m=4)			Maio (m=5)			Junho (m=6)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	160,96	168,14	1	133,95	141,13	1*	102,64	<b>109,83</b>
2	<b>148,34</b>	<b>157,92</b>	2	127,11	136,69	2	100,34	109,92
3	150,31	162,28	3	<b>117,00</b>	<b>128,98</b>	3	<b>99,31</b>	111,28
Julho (m=7)			Agosto (m=8)			Setembro (m=9)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	85,85	93,04	1	<b>90,63</b>	<b>97,81</b>	1	<b>161,32</b>	<b>168,50</b>
2	<b>71,37</b>	<b>80,96</b>	2	91,58	101,16	2	162,70	172,28
3	73,06	85,03	3	92,56	104,54	3	164,07	176,04
Outubro (m=10)			Novembro (m=11)			Dezembro (m=12)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	196,99	204,18	1*	189,03	<b>196,22</b>	1	<b>201,97</b>	<b>209,15</b>
2***	192,01	201,59	2	<b>188,73</b>	198,31	2	202,34	211,91
3	<b>186,87</b>	<b>198,85</b>	3	190,44	202,41	3	203,75	215,72

obtemos a ordem 2.

As estimativas de máxima verossimilhança e os intervalos de confiança dos parâmetros para as ordens escolhidas são apresentados na Tabela 5.2, observamos que todos os parâmetros são significativos. Note também que para os meses de fevereiro e agosto a estimativa pontual do parâmetro  $s$  ficou muito próxima do valor 2, para os meses janeiro, outubro e novembro a estimativa de  $s$  ficou acima de 2, indicando caudas mais leves que as da distribuição normal. Para os demais meses o valor de  $s$  indicou caudas mais pesadas que as da distribuição normal.

Analisando as estimativas intervalares para o parâmetro  $s$  observamos que o valor 2 só não está contido nos intervalos de confiança dos meses de junho e julho, indicando assim que a distribuição normal pode ser utilizada nos demais meses. Mas a fim de comparação ajustamos o modelo periódico autorregressivo para todos os meses utilizando a distribuição normal. Apresentamos os resultados dos critérios AIC e BIC na Tabela 5.3, já as estimativas de máxima verossimilhança dos parâmetros para as ordens escolhidas são apresentadas na Tabela 5.4.



Tabela 5.2: Estimativas de máxima verossimilhança utilizando a distribuição log-normal generalizada

	EMV	IC	EMV	IC	EMV	IC
	Janeiro (m=1)		Fevereiro (m=2)		Março (m=3)	
$\phi_1$	0,5473	(0,4010; 0,6937)	0,5032	(0,3133; 0,6931)	0,6123	(0,4364; 0,7881)
$\sigma_m^*$	1,4636	(1,1832; 1,7440)	1,2595	(0,8422; 1,6768)	1,0380	(0,7056; 1,3704)
s	3,6553	(1,0838; 6,2268)	2,1185	(0,8050; 3,4320)	1,7691	(0,9831; 2,5552)
	Abril (m=4)		Maio (m=5)		Junho (m=6)	
$\phi_1$	0,6021	(0,4779; 0,7262)	0,5147	(0,3339; 0,6955)	0,8427	(0,7771; 0,9084)
$\phi_2$	0,2750	(0,1493; 0,4006)	0,2149	(0,0524; 0,3773)		
$\phi_3$			0,2592	(0,1258; 0,3926)		
$\sigma_m^*$	0,6373	(0,3656; 0,9090)	0,5782	(0,3428; 0,8137)	0,4395	(0,2620; 0,6170)
s	1,3966	(0,7826; 2,0106)	1,6016	(0,7761; 2,4270)	1,2541	(0,8203; 1,6880)
	Julho (m=7)		Agosto (m=8)		Setembro (m=9)	
$\phi_1$	0,6022	(0,3676; 0,8368)	0,9104	(0,8214; 0,9995)	0,7446	(0,5875; 0,9017)
$\phi_2$	0,3481	(0,0099; 0,6863)				
$\sigma_m^*$	0,3756	(0,2237; 0,5275)	0,5728	(0,3847; 0,7608)	0,7999	(0,5033; 1,0966)
s	1,3169	(0,8231; 1,8106)	1,9722	(0,8930; 3,0515)	1,6631	(0,8530; 2,4733)
	Outubro (m=10)		Novembro (m=11)		Dezembro (m=12)	
$\phi_1$	0,3243	(0,0791; 0,5695)	0,6471	(0,4771; 0,8171)	0,5835	(0,3987; 0,7683)
$\phi_2$	0,3374	(0,0942; 0,5806)				
$\sigma_m^*$	1,1645	(0,9026; 1,4265)	1,1551	(0,8718; 1,4384)	0,9512	(0,5825; 1,3199)
s	2,4352	(1,3333; 3,5370)	2,4225	(1,1803; 3,6647)	1,4959	(0,8448; 2,1469)

Comparando as ordens escolhidas através dos ajustes com as distribuições log-normal generalizada e log-normal observamos que a escolha só diferiu no mês de junho.

Comparando as estimativas de máxima verossimilhança dos parâmetros  $\phi$ 's quando utilizamos as distribuições log-normal e log-normal generalizada observamos que as estimativas estão próximas, exceto para o mês de junho, o que já era esperado pois foram escolhidas ordens diferentes para esse mês.

O valor dos critérios AIC e BIC, quando utilizamos as distribuições log-normal e log-normal generalizada é a soma dos valores desses critérios para cada mês. O resultado é apresentado na Tabela 5.5. Através da mesma, notamos que o modelo utilizando a distribuição log-normal apresenta menor valor de ambos os critérios, o que já poderia ser esperado, pois em apenas dois meses (junho e julho) o parâmetro adicional, comparando

Tabela 5.3: Seleção dos modelos utilizando os critérios AIC e BIC quando adotamos a distribuição log-normal

Janeiro (m=1)			Fevereiro (m=2)			Março (m=3)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>208,39</b>	<b>213,18</b>	1***	210,50	215,29	1*	195,34	<b>200,12</b>
2	210,36	217,54	2	<b>206,93</b>	<b>214,12</b>	2	<b>194,23</b>	201,41
3	211,91	221,49	3	208,74	218,32	3	198,30	207,87
Abril (m=4)			Maio (m=5)			Junho (m=6)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	159,83	164,62	1	132,08	136,87	1	108,61	113,40
2	<b>148,95</b>	<b>156,13</b>	2	125,20	132,38	2	<b>106,04</b>	<b>113,23</b>
3	150,63	160,21	3	<b>115,71</b>	<b>125,28</b>	3	106,30	115,87
Julho (m=7)			Agosto (m=8)			Setembro (m=9)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	86,66	91,45	1	<b>88,63</b>	<b>93,42</b>	1	<b>159,86</b>	<b>164,65</b>
2	<b>74,78</b>	<b>81,97</b>	2	89,59	96,77	2	160,72	167,91
3	75,98	85,56	3	90,57	100,15	3	162,19	171,77
Outubro (m=10)			Novembro (m=11)			Dezembro (m=12)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	195,33	200,12	1*	187,57	<b>192,36</b>	1*	201,80	<b>206,59</b>
2***	190,76	197,94	2	<b>186,75</b>	193,93	2	<b>201,75</b>	208,93
3	<b>185,31</b>	<b>194,89</b>	3	188,46	198,04	3	203,67	213,24

Tabela 5.4: Estimativas de máxima verossimilhança utilizando a distribuição log-normal

	EMV	IC	EMV	IC	EMV	IC
Janeiro (m=1)		Fevereiro (m=2)		Março (m=3)		
$\phi_1$	0,5127	(0,3229; 0,7025)	0,5014	(0,3113; 0,6914)	0,6041	(0,4321; 0,7760)
$\sigma_m^*$	0,8545	(0,7229; 0,9860)	0,8657	(0,7324; 0,9990)	0,7883	(0,6669; 0,9097)
Abril (m=4)		Maio (m=5)		Junho (m=6)		
$\phi_1$	0,5868	(0,4253; 0,7484)	0,5194	(0,3442; 0,6946)	0,7117	(0,5285; 0,8948)
$\phi_2$	0,3062	(0,1456; 0,4668)	0,2123	(0,0466; 0,3780)	0,2035	(0,0195; 0,3874)
$\phi_3$			0,2508	(0,1109; 0,3907)		
$\sigma_m^*$	0,5848	(0,4947; 0,6748)	0,4704	(0,3980; 0,5429)	0,4487	(0,3796; 0,5178)
Julho (m=7)		Agosto (m=8)		Setembro (m=9)		
$\phi_1$	0,6114	(0,4369; 0,7860)	0,9107	(0,8218; 0,9995)	0,7657	(0,6275; 0,9039)
$\phi_2$	0,3452	(0,1713; 0,5192)				
$\sigma_m^*$	0,3700	(0,3130; 0,4269)	0,4080	(0,3452; 0,4708)	0,6333	(0,5358; 0,7308)
Outubro (m=10)		Novembro (m=11)		Dezembro (m=12)		
$\phi_1$	0,3202	(0,0599; 0,5805)	0,6608	(0,4939; 0,8277)	0,5509	(0,3708; 0,7309)
$\phi_2$	0,3455	(0,0866; 0,6044)				
$\sigma_m^*$	0,7570	(0,6404; 0,8735)	0,7514	(0,6357; 0,8671)	0,8204	(0,6941; 0,9468)

as duas distribuições, foi significativamente diferente de 2, o que implica em um ganho no ajuste utilizando a distribuição log-normal generalizada. Para os demais meses a utilização da distribuição log-normal generalizada acarretou em uma maior penalidade, pois apresenta um parâmetro a mais, sem um ganho expressivo no ajuste.

Tabela 5.5: Resultado dos critérios AIC e BIC dos ajustes utilizando os modelos LNG-PAR e LN-PAR

	AIC	BIC
LN-PAR	<b>1888,33</b>	<b>1960,16</b>
LNG-PAR	1891,39	1989,59

Adotamos que o melhor modelo é uma mescla entre as distribuições log-normal e log-normal generalizada, ou seja, escolhemos a ordem e a distribuição que melhor se ajusta a cada mês, com isso o resultado é a utilização da distribuição log-normal generalizada nos meses de junho e julho, para os demais meses a distribuição log-normal é a escolhida. Para o mês de janeiro e abril, o critério AIC com a distribuição log-normal generalizada é menor do que com a distribuição log-normal, mas esse resultado se inverte quando comparamos o critério BIC, escolhemos então pela distribuição com menor quantidade de parâmetros. Portanto, a ordem do modelo autorregressivo periódico final é  $(1,1,1,2,3,1,2,1,1,2,1,1)$ .

A partir das estimativas de máxima verossimilhança do modelo final calculamos os resíduos. Para verificar a qualidade do ajuste analisamos as funções de autocorrelação periódica e autocorrelação parcial periódica dos resíduos e as funções de autocorrelação e autocorrelação parcial dos resíduos agrupados. Constatamos que o ajuste eliminou a correlação existente na série.

As previsões para a série original são apresentadas na Figura 5.2, isso para os dois cenários já adotados anteriormente. As previsões podem ser comparadas através da Figura 3.4. Observamos através desses gráficos pequenas diferenças em ambos os cenários.

Através da Tabela 5.6 apresentamos os resultados dos critérios MAPE, EQM e EAM que medem a eficiência das previsões para o ano de 2012 da série de vazões de Furnas quando utilizamos o modelo autorregressivo periódico com mescla na distribuição condicional.

Comparando os resultados das métricas MAPE, EQM e EAM obtidas a partir

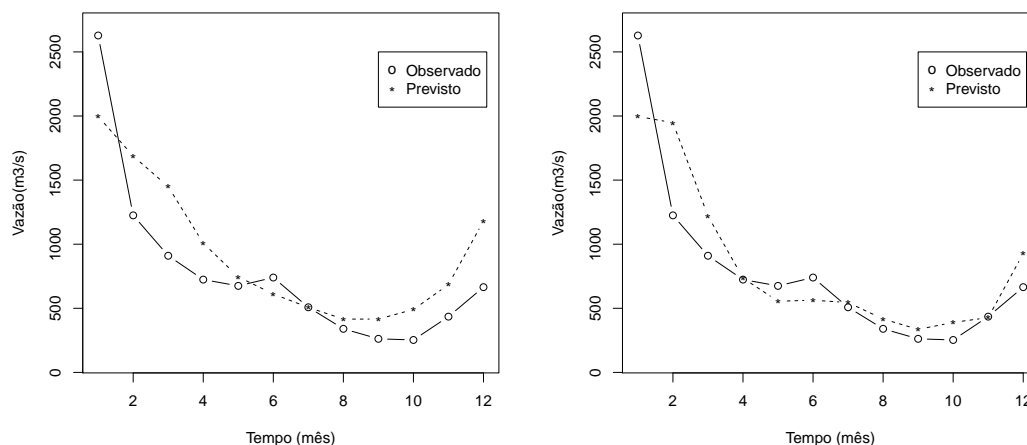


Figura 5.2: Previsões para a série original utilizando o modelo autorregressivo periódico com mescla na distribuição condicional

Tabela 5.6: Resumos dos critérios MAPE, EQM e EAM para as previsões da série de Furnas utilizando o modelo autorregressivo periódico com mescla na distribuição condicional

Cenário	MAPE(%)	EQM	EAM
1º	41,5	118.344,7	279,2
2º	26,1	96.989,0	214,0

das predições calculadas dos modelos PAR com a mescla das distribuições log-normal generalizada e log-normal e com o LNG-ARMA, apresentadas nas Tabelas 5.6 e 3.3, respectivamente, notamos que em ambos os cenários as predições utilizando o modelo PAR com a mescla das distribuições log-normal generalizada e log-normal alcançaram melhores resultados nas três métricas utilizadas.

## 5.2 Modelo Box-Cox Normal Generalizada Autorregressivo Periódico

A utilização da transformação de Box-Cox é bastante aplicada em estudos de séries temporais periódicas, principalmente quando a série em estudo é um série de vazões, como pode ser visto no livro do Hipel e McLeod (1994). A transformação é utilizada para a correção de assimetria e estabilizar a variabilidade da série original.

Considere que  $\{X_{t(r,m)}; r \geq 1; m = 1, \dots, S\}$  seja uma série temporal periódica, empregando a transformação de Box-Cox na série original temos

$$Y_{t(r,m)}^{(\lambda)} = \begin{cases} \frac{X_{t(r,m)}^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(X_{t(r,m)}), & \lambda = 0 \end{cases} \quad (5.3)$$

sendo  $\{Y_{t(r,m)}^{(\lambda)}; r \geq 1; m = 1, \dots, S\}$  a série transformada. Como definimos que a série original apresenta periodicidade então optamos por retirar a sazonalidade da série transformada utilizando a equação dada por

$$Z_{t(r,m)}^{(\lambda)} = \frac{Y_{t(r,m)}^{(\lambda)} - \mu_m}{\sigma_m}, \quad (5.4)$$

sendo que  $\mu_m = E\{Y_{t(r,m)}^{(\lambda)}\}$ ,  $\sigma_m = \sqrt{\text{Var}\{Y_{t(r,m)}^{(\lambda)}\}}$  com  $r \geq 1$  e  $m = 1, 2, \dots, S$  e  $\{Z_{t(r,m)}^{(\lambda)}; r \geq 1; m = 1, \dots, S\}$  é a série resultante.

Assumimos que a distribuição condicional da série temporal transformada e livre de sazonalidade ainda tem comportamento periódico, ou seja, apresenta função de autocorrelação periódica, por isso atribuímos que a distribuição condicional da série resultante para todo instante de tempo tem distribuição normal generalizada com os parâmetros  $\mu_m^*$ ,  $\sigma_m^*$  e  $s_m$ , ou seja,

$$Z_{t(r,m)}^{(\lambda)} | \mathfrak{F}_{t(r,m)-1} \sim \text{NG}(\mu_m^*, \sigma_m^*, s_m), \quad (5.5)$$

sendo que  $\mu_m^*$  é dado pela equação (4.3).

### 5.2.1 Estimação de Máxima Verossimilhança

Denotamos os parâmetros do modelo BC-NG-PAR a serem estimados por  $(\lambda, \phi_1, \dots, \phi_S, \sigma, \mathbf{s})$  sendo que  $\phi_m = \{\phi_{1,m}, \dots, \phi_{p_m,m}\}$  com  $m = 1, \dots, S$ ,  $\sigma = \{\sigma_1^*, \dots, \sigma_m^*\}$  e  $\mathbf{s} = \{s_1, \dots, s_m\}$ . A estimação de máxima verossimilhança será feita em dois passos, de maneira parecida à desenvolvida no Capítulo 3. Definimos o parâmetro  $\lambda$  como perturbador e o restante como parâmetros de interesse. Sejam  $\mathbf{x} = \{x_1, \dots, x_n\}$  a série temporal observada e  $\{z_{t(r,m)}^{(\lambda)}; r = 1, \dots, R, m = 1, \dots, S\}$  a série transformada e livre de sazonalidade, para algum valor de  $\lambda$  em (5.3). Para cada  $m$  a função verossimilhança, condicional nas primeiras observações e ao valor de  $\lambda$ , é dada por

$$L(\phi_m, \sigma_m^*, s_m | \lambda, \mathbf{x}) = \prod_{r=r_{1,m}}^R f(z_{t(r,m)} | \mathfrak{F}_{t(r,m)-1}) J_{ZX},$$

sendo  $r_{1,m} = 1$  se  $p_m < m$  ou  $r_{1,m} = 2$  se  $m \leq p_m < S$  e  $J_{ZX} = \frac{\partial Z}{\partial X}$ . Se  $\lambda \neq 0$  observe que

$$z_{t(r,m)}^{(\lambda)} = \frac{x_{t(r,m)}^{\lambda-1} - \mu_m}{\sigma_m}$$

então

$$L(\phi_m, \sigma_m^*, s_m | \lambda, \mathbf{x}) = \prod_{r=r_{1,m}}^R \frac{x_{t(r,m)}^{\lambda-1}}{\sigma_m} \left( \frac{s}{2\sigma_m^* \Gamma(1/s)} \right)^{R-r_{1,m}} \times \exp \left\{ -\frac{1}{(\sigma_m^*)^s} \sum_{r=r_{1,m}}^R \left| z_{t(r,m)}^{(\lambda)} - \sum_{m=1}^{p_m} \phi_{j,m} z_{t(r,m)-j}^{(\lambda)} \right|^s \right\}, \quad (5.6)$$

mas caso  $\lambda = 0$  então

$$z_{t(r,m)}^{(\lambda)} = \frac{\log(x_{t(r,m)}) - \mu_m}{\sigma_m},$$

logo

$$L(\phi_m, \sigma_m^*, s_m | \lambda, \mathbf{x}) = \prod_{r=r_{1,m}}^R \frac{1}{x_{t(r,m)} \sigma_m} \left( \frac{s}{2\sigma_m^* \Gamma(1/s)} \right)^{R-r_{1,m}} \times \exp \left\{ -\frac{1}{(\sigma_m^*)^s} \sum_{r=r_{1,m}}^R \left| z_{t(r,m)}^{(\lambda)} - \sum_{m=1}^{p_m} \phi_{j,m} z_{t(r,m)-j}^{(\lambda)} \right|^s \right\}, \quad (5.7)$$

note que nas equações (5.6) e (5.7) temos  $\sigma_m^*$  e  $\sigma_m$ , sendo que  $\sigma_m$  é estimado através da equação  $\sigma_m = \sqrt{\text{Var}\{Y_{t(r,m)}^{(\lambda)}\}}$  enquanto que  $\sigma_m^*$  é estimado a partir da função de verossimilhança. Observe também que a obtenção das estimativas de máxima verossimilhança dos parâmetros para cada  $m$  podem ser encontradas separadamente.

Notemos que a equação de máxima verossimilhança condicional em  $\lambda$  é bastante complexa, por isso indicamos e utilizamos métodos iterativos para a obtenção das estimativas de máxima verossimilhança, por exemplo, o método “L-BFGS-B” que está disponível no *softwares* R Core Team (2014) (Byrd *et al.*, 1995). Em muitos *software* a função log-verossimilhança é requerida. Para  $\lambda \neq 0$  a função log-verossimilhança é dada por

$$\begin{aligned} l(\phi_m, \sigma_m^*, s_m | \lambda, \mathbf{x}) &= \log[L(\phi_m, \sigma_m^*, s_m | \lambda, \mathbf{x})] \\ &= (\lambda - 1) \sum_{r=r_{1,m}}^R \log(x_{t(r,m)}) - R \log(\sigma_m) \\ &\quad + (R - r_{1,m}) \log \left( \frac{s}{2\sigma_m^* \Gamma(1/s)} \right) \\ &\quad - \frac{1}{(\sigma_m^*)^s} \sum_{r=r_{1,m}}^R \left| z_{t(r,m)}^{(\lambda)} - \sum_{m=1}^{p_m} \phi_{j,m} z_{t(r,m)-j}^{(\lambda)} \right|^s, \end{aligned}$$

e quando  $\lambda = 0$  a função log-verossimilhança é dada por

$$\begin{aligned} l(\phi_m, \sigma_m^*, s_m | \lambda, \mathbf{x}) &= - \sum_{r=r_{1,m}}^R \log(x_{t(r,m)}) - R \log(\sigma_m) \\ &\quad + (R - r_{1,m}) \log \left( \frac{s}{2\sigma_m^* \Gamma(1/s)} \right) \\ &\quad - \frac{1}{(\sigma_m^*)^s} \sum_{r=r_{1,m}}^R \left| z_{t(r,m)}^{(\lambda)} - \sum_{m=1}^{p_m} \phi_{j,m} z_{t(r,m)-j}^{(\lambda)} \right|^s. \end{aligned}$$

Os intervalos de confiança assintóticos são obtidos através das estimativas pontuais e das variâncias amostrais. Iniciamos a investigação da ordem do modelo BC-NG-PAR, para cada um dos períodos, através das funções de autocorrelação periódica e autocorrelação parcial periódica. Mas adotamos os critérios AIC e BIC para a escolha da ordem em cada um dos períodos, que são dados por

$$\begin{aligned} AIC_m &= -2l(\hat{\phi}_{1,m}, \dots, \hat{\phi}_{p_m,m}, \hat{\sigma}_m^*, \hat{s}_m | \lambda, \mathbf{z}) + 2(p_m + 2) \\ BIC_m &= -2l(\hat{\phi}_{1,m}, \dots, \hat{\phi}_{p_m,m}, \hat{\sigma}_m^*, \hat{s}_m | \lambda, \mathbf{z}) + (p_m + 2) \log(R - r_{1,m}), \end{aligned}$$

sendo que  $\hat{\phi}_{i,m}, \hat{\sigma}_m^*, \hat{s}_m$ , com  $i = 1, \dots, p_m$ , são as estimativas de máxima verossimilhança dos respectivos parâmetros, quando todos os parâmetros autorregressivos são significativos. Os valores dos critérios AIC e BIC resultante do ajuste são dados pela soma dos critérios para cada período.

Observamos que a função de máxima verossimilhança para cada  $m$  descrita anteriormente é condicional ao parâmetro  $\lambda$ . Obtemos a estimativa de máxima verossimilhança e o intervalo de confiança assintótico deste parâmetro a partir do modelo BC-NG-ARMA, como descrito na Seção 3.2.1. Destacamos que neste trabalho o valor do parâmetro  $\lambda$  é único para todos os  $m$ 's.

## 5.2.2 Previsão

O nosso interesse é a previsão de valores da série original, mas devido ao fato que ajustamos o modelo à série sem sazonalidade precisamos primeiramente prever valores para a série livre de sazonalidade, para depois calcular as previsões para a série original.

Adotamos que a série temporal foi observada até o instante  $t = t(r_0, m_0)$ , e supomos que queremos prever valores da série  $h$  passos a frente. Optamos pela previ-

são que minimiza o erro quadrático médio, ou seja, assumimos que o valor previsto é dado através da média. Definimos que a previsão de *origem*  $t$  e *horizonte*  $h$  é dada por  $\widehat{X}_{t(r_0, m_0)}(h) = \widehat{E}(X_{t(r_0, m_0)+h} | \mathfrak{F}_{t(r_0, m_0)})$ , note que essa previsão é para a série original.

A previsão a um passo a frente ( $h = 1$ ), para a série livre de sazonalidade, utilizando o modelo BC-NG-PAR é dada por

$$\widehat{Z}_{t(r_0, m_0)}^{(\lambda)}(1) = \sum_{j=1}^{p_{m_0+1}} \widehat{\phi}_{j, m_0+1} z_{(r_0-1)S+m_0+1-j}^{(\lambda)}, \quad (5.8)$$

sendo que  $\widehat{\phi}_{j, m_0+1}$ ,  $j = 1, \dots, p_{m_0+1}$  são as estimativas de máxima verossimilhança dos respectivos parâmetros. Note que a equação (5.8) é a equação (4.3) substituindo os parâmetros  $\phi_{j, m_0+1}$  por suas estimativas de máxima verossimilhança, portanto a equação de previsão a um passo a frente é dada através da equação da média da distribuição normal generalizada.

Note que precisamos do valor do instante  $t + 1$  para predizer o valor no instante  $t + 2$ , o que pode ser resolvido ou observando o valor no instante  $t + 1$  e fazendo a previsão a um passo a frente, ou considerando o valor predito no instante  $t + 1$  como sendo o valor observado e assim predizendo o instante  $t + 2$ .

Para o cálculo das previsões para *horizonte* maior que 1, quando não acrescentamos novas informações, definimos que

$$\widehat{E}[Z_{t(r_0, m_0)+h-j}^{(\lambda)} | \mathfrak{F}_t] = \begin{cases} z_{t(r_0, m_0)+h-j}^{(\lambda)} & \text{se } h \leq j \\ \widehat{Z}_{t(r_0, m_0)}^{(\lambda)}(h-j) & \text{se } h > j, \end{cases} \quad (5.9)$$

com isso a previsão para  $h$  passos a frente é dada por

$$\begin{aligned} \widehat{Z}_{t(r_0, m_0)}^{(\lambda)}(h) &= \widehat{E}(Z_{t(r_0, m_0)+h}^{(\lambda)} | \mathfrak{F}_t) \\ &= \widehat{E}(Z_{t(r_1, m_1)}^{(\lambda)} | \mathfrak{F}_t) \\ &= \sum_{j=1}^{p_{m_1}} \widehat{\phi}_{j, m_1} \widehat{E}[Z_{t(r_1, m_1)-j}^{(\lambda)} | \mathfrak{F}_t], \end{aligned} \quad (5.10)$$

em que  $(r_0 - 1)S + m_0 + h = (r_1 - 1)S + m_1$  e  $\widehat{\phi}_{j, m_1}$ , com  $j = 1, \dots, p_{m_1}$ , são as estimativas de máxima verossimilhança dos respectivos parâmetros.

A partir das equações (5.8) e (5.10) conseguimos predizer a série livre de sazonalidade para qualquer *horizonte*. Como o nosso interesse é a obtenção das previsões para



a série original, observe que

$$P(Z_{t(r,m)+h} \leq z_{t(r,m)+h}^{(M)} | \mathfrak{F}_t) = 0,5,$$

assim  $z_{t(r,m)+h}^{(M)}$  é o valor da mediana da distribuição  $Z_{t+h} | \mathfrak{F}_t$ , mas conseguimos escrever a série  $Z_t$  como função da série  $X_t$ , disto temos que

$$P\left(\frac{(X_{t(r_1, m_1)}^{(M)})^{\hat{\lambda}} - \hat{\mu}_{m_1}}{\hat{\sigma}_{m_1}} \leq z_{t+h}^{(M)} | \mathfrak{F}_t\right) = 0,5,$$

sendo  $t(r_0, m_0) + h = t(r_1, m_1)$ . Definindo  $\mu_t^*(h)$  como a previsão para  $X_{t(r,m)+h}$  temos que

$$\mu_t^*(h) = [\hat{\lambda}(z_{t+h}^{(M)} \hat{\sigma}_{m_1} + \hat{\mu}_{m_1}) + 1]^{1/\hat{\lambda}},$$

como a distribuição normal generalizada é simétrica em torno da média, e a média é igual a mediana, podemos aproximar a mediana pela previsão da série sem sazonalidade, resultando em

$$\mu_t^*(h) = [\hat{\lambda}(\hat{\mu}_{t+h} \hat{\sigma}_{m_1} + \hat{\mu}_{m_1}) + 1]^{1/\hat{\lambda}},$$

para qualquer  $h \geq 1$ . Portanto o cálculo das previsões para a série original depende das previsões para a série sem sazonalidade.

### 5.2.3 Estudo de Simulação

Primeiramente vamos esquematizar o processo de geração de séries temporais que seguem o modelo BC-NG-PAR. Os passos a seguir são utilizados para a geração de uma série temporal:

1. definir os valores dos parâmetros  $S, R, p_1, \dots, p_S, \phi_1, \dots, \phi_m, \sigma, s$ ;
2. definir os valores iniciais de  $z_{t(r,m)}$ ;
3. calcular  $\mu_{t(r,m)}^* = \sum_{j=1}^{p_m} \phi_{j,m} z_{t(r,m)-j}$ ,
4. gerar  $z_{t(r,m)}$  da distribuição  $\text{NG}(\mu_{t(r,m)}^*, \sigma_m^*, s_m)$ ,
5. calcular  $x_{t(r,m)} = (\lambda z_{t(r,m)} + 1)^{1/\lambda}$ ;

6. repetir os itens 3 a 5 até conseguir uma série com  $K$  valores a mais do que se deseja;
7. desconsiderar os  $K$  primeiros valores da série para conseguir o tamanho da série desejada.

A geração esquematizada leva em consideração que  $\lambda \neq 0$ , mas caso o interesse seja a geração de séries para  $\lambda = 0$  então basta substituir a equação do item 5 por  $x_{t(r,m)} = \exp(y_{t(r,m)})$ . O descarte dos  $K$  primeiros valores da série gerada é para diminuir a influência dos valores iniciais.

Neste estudo de simulação estamos interessados em analisar o comportamento das estimativas de máxima verossimilhança com respeito às métricas vício absoluto médio (VAM), o erro quadrático médio (EQM), a média das estimativas (ME) e a probabilidade de cobertura (PC), para cada um dos parâmetros. Para isto repetimos 1000 vezes o processo de geração e estimação, considerando  $S = 12$  e adotando  $R = 30, 50, 80$  e  $100$ , com isso estamos considerando séries com tamanho igual a  $360, 600, 960$  e  $1200$ . Em cada uma das repetições o interesse é nas estimativas de máxima verossimilhança e na matriz de informação observada de Fisher.

Para este estudo adotamos que o verdadeiro valor de  $\lambda$  é 0 e dos demais parâmetros são os apresentados na Tabela 5.7. Notemos que no vetor  $\mathbf{s}$  existem valores maiores e menores que 2, assim conseguimos gerar séries com caudas mais leves e mais pesadas do que as da distribuição normal. Apresentamos os resultados das métricas ME, VAM, EQM e PC nas Tabelas 5.8, 5.9, 5.10 e 5.11.

Tabela 5.7: Parâmetros adotados para o estudo de simulação do modelo NG-PAR para a análise das métricas ME, VAM, EQM e PC

	m											
	1	2	3	4	5	6	7	8	9	10	11	12
$\phi_{1,m}$	0,60	0,50	0,40	0,35	0,30	0,70	0,80	0,75	0,60	0,50	0,30	0,55
$\sigma_m^*$	1,10	1,50	0,90	0,85	1,20	1,00	1,80	1,30	0,60	0,50	1,40	0,75
$s_m$	1,60	1,50	2,40	1,70	2,00	1,40	1,80	2,50	1,90	1,75	1,60	2,30

Observamos que a média das estimativas se apresentam próximas do verdadeiro valor do parâmetro mesmo com séries de tamanho pequeno. As métricas vício absoluto médio e erro quadrático médio diminuem com o aumento do tamanho da série. A

Tabela 5.8: Estudo de simulação do modelo BC-NG-PAR considerando as métricas VAM, ME, EQM e PC  $m = 1, 2$  e  $3$  quando o valor verdadeiro de  $\lambda$  é  $0$

	VAM	ME	EQM	PC	VAM	ME	EQM	PC	VAM	ME	EQM	PC
	n=360				n=360				n=360			
$\phi$	0,1770	0,5999	0,0508	0,7590	0,1883	0,4972	0,0591	0,7400	0,0702	0,3973	0,0079	0,8780
$\sigma_m^*$	0,3056	1,0857	0,1425	0,8720	0,4464	1,4921	0,3066	0,8600	0,1495	0,8672	0,0377	0,9240
$s$	0,9689	2,1940	3,0606	0,9240	0,9063	2,0597	2,7775	0,9200	1,4077	3,2939	5,6502	0,9420
	n=600				n=600				n=600			
$\phi$	0,1311	0,5980	0,0275	0,8600	0,1475	0,4977	0,0332	0,8130	0,0488	0,3980	0,0039	0,9150
$\sigma_m^*$	0,2308	1,1242	0,0846	0,8880	0,3363	1,5173	0,1798	0,8910	0,1145	0,8938	0,0213	0,9230
$s$	0,6701	2,0071	1,5222	0,9450	0,5809	1,8281	1,1602	0,9340	1,1869	3,1870	4,2253	0,9540
	n=960				n=960				n=960			
$\phi$	0,1083	0,6038	0,0183	0,8700	0,1103	0,4983	0,0196	0,8520	0,0404	0,4010	0,0026	0,9250
$\sigma_m^*$	0,1767	1,0973	0,0486	0,9240	0,2551	1,4944	0,1001	0,9200	0,0932	0,8989	0,0138	0,9320
$s$	0,3947	1,7539	0,3181	0,9460	0,3532	1,6279	0,2490	0,9340	0,7700	2,8427	1,8228	0,9640
	n=1200				n=1200				n=1200			
$\phi$	0,0876	0,5978	0,0122	0,9230	0,1025	0,5036	0,0163	0,8900	0,0337	0,4001	0,0018	0,9370
$\sigma_m^*$	0,1537	1,0993	0,0373	0,9330	0,2407	1,5150	0,0913	0,9040	0,0876	0,8958	0,0117	0,9440
$s$	0,3375	1,7276	0,3025	0,9610	0,3201	1,6276	0,2241	0,9510	0,5958	2,6697	0,7008	0,9600

Tabela 5.9: Estudo de simulação do modelo NG-PAR considerando as métricas VAM, ME, EQM e PC para  $m = 4, 5$  e  $6$  quando o valor verdadeiro de  $\lambda$  é  $0$

	m=4				m=5				m=6			
	VAM	ME	EQM	PC	VAM	ME	EQM	PC	VAM	ME	EQM	PC
	n=360				n=360				n=360			
$\phi$	0,1342	0,3437	0,0285	0,7900	0,1342	0,2990	0,0596	0,8570	0,1585	0,7015	0,0421	0,7060
$\sigma_m^*$	0,2080	0,8329	0,0691	0,8870	0,2080	1,1855	0,0913	0,9010	0,3028	0,9932	0,1411	0,8780
$s$	0,9547	2,2830	3,0160	0,9180	0,9547	2,7930	4,1792	0,9490	0,7711	1,8543	2,2183	0,9310
	n=600				n=600				n=600			
$\phi$	0,0994	0,3518	0,0156	0,8750	0,0994	0,2972	0,0300	0,9120	0,1213	0,6969	0,0235	0,7940
$\sigma_m^*$	0,1643	0,8573	0,0413	0,9060	0,1643	1,2041	0,0590	0,8940	0,2279	1,0189	0,0813	0,9140
$s$	0,7117	2,1269	1,5126	0,9480	0,7117	2,5784	2,5597	0,9490	0,4815	1,6564	0,6866	0,9470
	n=960				n=960				n=960			
$\phi$	0,0744	0,3573	0,0087	0,9220	0,0744	0,3039	0,0206	0,9110	0,0969	0,6948	0,0149	0,8050
$\sigma_m^*$	0,1286	0,8660	0,0256	0,9150	0,1286	1,2000	0,0357	0,9170	0,1849	0,9951	0,0536	0,9110
$s$	0,4595	1,9457	0,5164	0,9470	0,4595	2,2944	1,0611	0,9580	0,3171	1,5052	0,2135	0,9390
	n=1200				n=1200				n=1200			
$\phi$	0,0667	0,3470	0,0070	0,9240	0,0667	0,2997	0,0149	0,9400	0,0848	0,7015	0,0113	0,8460
$\sigma_m^*$	0,1108	0,8536	0,0192	0,9360	0,1108	1,2007	0,0285	0,9390	0,1647	1,0039	0,0417	0,9290
$s$	0,3577	1,8440	0,2604	0,9520	0,3577	2,2029	0,5440	0,9560	0,2786	1,5031	0,1543	0,9440

probabilidade de cobertura, em geral, fica abaixo do nível nominal para séries de tamanho pequeno, mas com o aumento do tamanho da série a probabilidade de cobertura se aproxima da nominal.

Verificamos um bom comportamento das métricas estudadas, indicando assim que a estimação do parâmetro  $\lambda$ , a partir dos dados, não influenciou o desempenho destas métricas.

Apresentamos na Figura 5.3 os histogramas das estimativas do parâmetro  $\lambda$  considerando os quatro tamanhos de séries. Por meio dela verificamos que a distribuição do parâmetro torna-se mais compacta com o aumento do tamanho da série, sendo que o

Tabela 5.10: Estudo de simulação do modelo NG-PAR considerando as métricas VAM, ME, EQM e PC para  $m = 7, 8$  e  $9$  quando o valor verdadeiro de  $\lambda$  é  $0$ 

	m=7				m=8				m=9			
	VAM	ME	EQM	PC	VAM	ME	EQM	PC	VAM	ME	EQM	PC
	n=360				n=360				n=360			
$\phi$	0,1996	0,8044	0,0673	0,8240	0,0813	0,7493	0,0105	0,8840	0,0471	0,6005	0,0035	0,8550
$\sigma_m^*$	0,4041	1,7761	0,2597	0,9000	0,2099	1,2470	0,0748	0,9130	0,1357	0,5823	0,0306	0,8950
$s$	1,1116	2,5568	4,3229	0,9290	1,4689	3,4028	6,3026	0,9320	1,1296	2,5924	4,2735	0,9070
	n=600				n=600				n=600			
$\phi$	0,1506	0,7946	0,0368	0,8770	0,0627	0,7496	0,0063	0,9190	0,0337	0,5973	0,0018	0,9220
$\sigma_m^*$	0,3193	1,7963	0,1655	0,9070	0,1634	1,2942	0,0424	0,9360	0,0975	0,6071	0,0152	0,9250
$s$	0,7203	2,2168	1,5172	0,9490	1,1315	3,2140	3,5735	0,9530	0,7721	2,3892	1,8563	0,9680
	n=960				n=960				n=960			
$\phi$	0,1173	0,8065	0,0215	0,8990	0,0498	0,7485	0,0040	0,9120	0,0275	0,6004	0,0012	0,9230
$\sigma_m^*$	0,2666	1,8135	0,1098	0,9150	0,1343	1,3008	0,0282	0,9300	0,0789	0,6003	0,0101	0,9190
$s$	0,5080	2,0489	0,9462	0,9600	0,7857	2,9405	1,5865	0,9560	0,5402	2,1731	1,0211	0,9440
	n=1200				n=1200				n=1200			
$\phi$	0,1051	0,8056	0,0170	0,9220	0,0403	0,7504	0,0026	0,9430	0,0249	0,5994	0,0010	0,9360
$\sigma_m^*$	0,2211	1,8209	0,0785	0,9280	0,1144	1,2946	0,0212	0,9410	0,0710	0,6070	0,0080	0,9260
$s$	0,4059	1,9905	0,3914	0,9620	0,6581	2,8086	1,0849	0,9500	0,4505	2,1240	0,5531	0,9590

Tabela 5.11: Estudo de simulação do modelo NG-PAR considerando as métricas VAM, ME, EQM e PC para  $m = 10, 11$  e  $12$  quando o valor verdadeiro de  $\lambda$  é  $0$ 

	m=10				m=11				m=12			
	VAM	ME	EQM	PC	VAM	ME	EQM	PC	VAM	ME	EQM	PC
	n=360				n=360				n=360			
$\phi$	0,0617	0,5052	0,0061	0,8360	0,3012	0,3094	0,1455	0,7750	0,0706	0,5478	0,0078	0,8770
$\sigma_m^*$	0,1199	0,4885	0,0229	0,8790	0,3607	1,3965	0,2038	0,8960	0,1299	0,7265	0,0278	0,9200
$s$	1,0148	2,3777	2,8410	0,9230	0,8627	2,1363	2,3141	0,9340	1,3407	3,1359	5,6732	0,9500
	n=600				n=600				n=600			
$\phi$	0,0484	0,5032	0,0038	0,8750	0,2123	0,2996	0,0726	0,8560	0,0523	0,5502	0,0043	0,9160
$\sigma_m^*$	0,0951	0,5028	0,0142	0,8940	0,2887	1,4268	0,1273	0,9010	0,0995	0,7542	0,0156	0,9210
$s$	0,7726	2,2182	2,2901	0,9520	0,7048	2,0427	1,8031	0,9520	1,0886	3,0322	4,2268	0,9620
	n=960				n=960				n=960			
$\phi$	0,0343	0,5011	0,0018	0,9230	0,1663	0,2914	0,0431	0,9060	0,0388	0,5517	0,0024	0,9400
$\sigma_m^*$	0,0737	0,4986	0,0085	0,9130	0,2195	1,4245	0,0741	0,9310	0,0803	0,7509	0,0100	0,9360
$s$	0,4775	1,9668	0,5002	0,9500	0,4313	1,8189	0,4870	0,9670	0,7197	2,7047	1,4638	0,9620
	n=1200				n=1200				n=1200			
$\phi$	0,0323	0,4986	0,0016	0,9280	0,1511	0,3129	0,0355	0,9240	0,0361	0,5497	0,0021	0,9280
$\sigma_m^*$	0,0637	0,5054	0,0065	0,9110	0,1960	1,4178	0,0599	0,9290	0,0756	0,7498	0,0092	0,9200
$s$	0,3976	1,9444	0,4280	0,9590	0,3442	1,7447	0,2414	0,9630	0,6240	2,6147	0,9543	0,9490

centro da distribuição é o verdadeiro valor do parâmetro.

## 5.2.4 Aplicação à Série de Furnas

Adotamos a série de vazões médias mensais afluente da usina hidrelétrica de Furnas apresentada na Seção 2.3.1. Como descrito na Seção 5.2.1 precisamos inicialmente estimar o valor de  $\lambda$  utilizando o modelo BC-NG-ARMA, mas na Seção 3.2.4 são apresentadas as estimativas pontuais  $\hat{\lambda} = -0.1$  e  $\hat{\lambda} = 0$ , e as intervalares  $\{-0, 2; -0, 1; 0\}$  e  $\{-0, 1; 0; 0, 1\}$ , respectivamente quando adotamos as distribuições normal e normal gene-

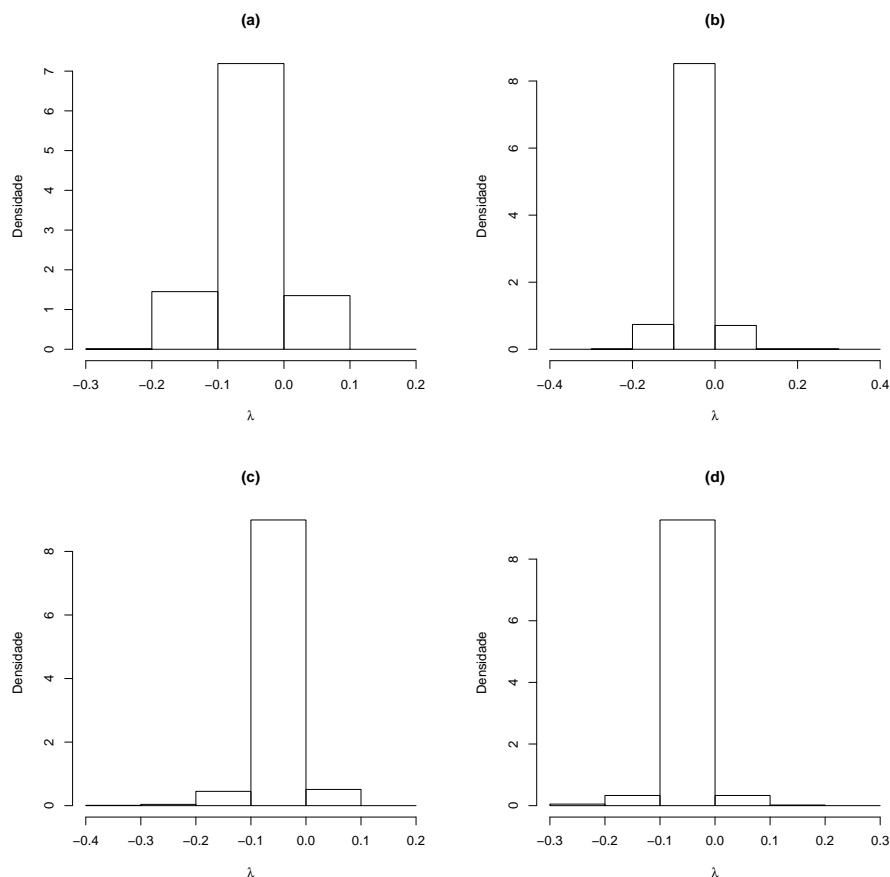


Figura 5.3: Histogramas da distribuição amostral de  $\hat{\lambda}$  quando ajustamos o modelo BC-NG-PAR adotando  $\lambda$  verdadeiro igual a 0,5 e tamanhos de séries iguais a 360, 600, 960 e 1200 em (a), (b), (c) e (d), respectivamente

realizada.

Utilizamos aqui a mesma estimativa de  $\lambda$  adotada na Seção 3.2.4, ou seja, a estimativa obtida por meio da distribuição normal, que é  $\hat{\lambda} = 0$ . Com isso transformamos os dados utilizando a função logarítmica, que é o mesmo se adotarmos o modelo LNG-PAR. Portanto o ajuste do modelo BC-NG-PAR à série de vazões médias mensais de Furnas se reduz ao ajuste do modelo LNG-PAR apresentado na Seção 5.1.2.

### 5.2.5 Aplicação à Série de Sobradinho

Consideramos agora a série de vazões médias mensais afluentes da usina hidrelétrica de Sobradinho, apresentada na Seção 3.2.5, na qual estimamos o valor de  $\lambda$  adotando as distribuições normal e normal generalizada, que são  $\hat{\lambda} = -0,1$  e  $\hat{\lambda} = -0,3$ , respectiva-

mente.

Utilizando a estimativa pontual de  $\lambda$ , quando adotamos a distribuição normal generalizada, encontramos as funções de autocorrelação periódica e autocorrelação parcial periódica, que são apresentadas na Figura 5.4. A partir dos gráficos notamos que a função de autocorrelação muda de um mês para o outro. As ordens indicadas por meio dos gráficos são (1,1,1,1,1,3,1,1,2,1,1,1), de janeiro a dezembro, respectivamente.

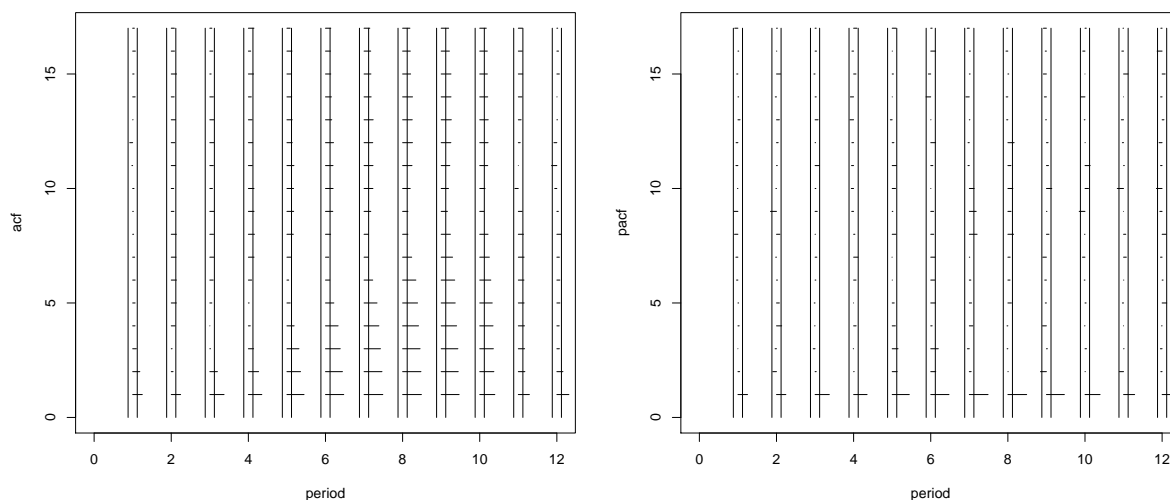


Figura 5.4: As funções de autocorrelação periódica (esquerda) e autocorrelação parcial periódica (direita) da transformação da série de sobradinho adotando  $\lambda = -0,3$

Escolhemos para cada mês as ordens do modelo utilizando os critérios AIC e BIC. Os resultados dos critérios de seleção considerando as distribuições normal generalizada e normal, são apresentados respectivamente nas Tabelas 5.12 e 5.13.

O significado dos símbolos \* e \*\*\* nas Tabelas 5.12 e 5.13 é o mesmo empregado anteriormente. Comparando as ordens escolhidas para cada mês utilizando as distribuições normal generalizada e normal, verificamos que apenas nos meses de maio e junho as ordens são diferentes. Nos demais meses, o ajuste contém a mesma quantidade de termos autorregressivos.

As estimativas de máxima verossimilhança para cada mês considerando as distribuições normal generalizada e normal, são apresentadas respectivamente nas Tabelas 5.14 e 5.15.

Tabela 5.12: Seleção da melhor ordem do modelo BC-NG-PAR utilizando os critérios AIC e BIC para explicar a série de Sobradinho quando adotamos  $\hat{\lambda} = -0.3$ 

Janeiro (m=1)			Fevereiro (m=2)			Março (m=3)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>-118,33</b>	<b>-111,72</b>	1	<b>-72,45</b>	<b>-65,83</b>	1	<b>-106,56</b>	<b>-99,94</b>
2	-118,16	-109,34	2	-67,87	-59,05	2	-104,57	-95,75
3	-116,20	-105,17	3	-66,19	-55,17	3	-99,20	-88,17
Abril (m=4)			Maio (m=5)			Junho (m=6)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>-92,42</b>	<b>-85,81</b>	1	-157,67	-151,06	1	-229,55	-222,94
2	-90,47	-81,65	2	-162,00	-153,18	2	-234,60	-225,78
3	-89,13	-78,10	3	<b>-172,19</b>	<b>-161,17</b>	3	<b>-242,24</b>	<b>-231,22</b>
Julho (m=7)			Agosto (m=8)			Setembro (m=9)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>-289,40</b>	<b>-282,79</b>	1*	-303,01	<b>-296,40</b>	1	-233,53	-226,91
2	-287,68	-278,86	2	-302,08	-293,26	2	<b>-240,89</b>	<b>-232,07</b>
3	-286,79	-275,77	3	<b>-307,13</b>	-296,10	3	-229,53	-218,50
Outubro (m=10)			Novembro (m=11)			Dezembro (m=12)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>-150,98</b>	<b>-144,37</b>	1*	-75,31	<b>-68,70</b>	1	<b>-91,79</b>	<b>-85,17</b>
2	-148,99	-140,17	2	<b>-76,02</b>	-67,20	2	-90,26	-81,44
3	-147,79	-136,76	3	-74,06	-63,03	3	-90,36	-79,34

Tabela 5.13: Seleção da melhor ordem do modelo BC-N-PAR utilizando os critérios AIC e BIC para explicar a série de Sobradinho quando adotamos  $\hat{\lambda} = -0.3$ 

Janeiro (m=1)			Fevereiro (m=2)			Março (m=3)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>-111,05</b>	<b>-106,64</b>	1	<b>-74,33</b>	<b>-69,92</b>	1	<b>-106,33</b>	<b>-101,92</b>
2	-109,30	-102,69	2	-69,69	-63,07	2	-104,59	-97,97
3	-107,30	-98,48	3	-68,11	-59,29	3	-99,15	-90,33
Abril (m=4)			Maio (m=5)			Junho (m=6)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1*	-80,60	<b>-86,19</b>	1	-156,56	-152,15	1	-230,88	-226,47
2	<b>-88,62</b>	-82,00	2***	-159,45	-152,83	2***	-233,95	-227,34
3	-88,33	-79,51	3	<b>-165,52</b>	<b>-156,70</b>	3	<b>-243,18</b>	<b>-234,36</b>
Julho (m=7)			Agosto (m=8)			Setembro (m=9)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>-290,47</b>	<b>-286,06</b>	1	<b>-296,14</b>	<b>-291,73</b>	1	-224,89	-220,48
2	-288,65	-282,03	2	-294,18	-287,56	2	<b>-229,88</b>	<b>-223,27</b>
3	-287,80	-278,98	3	-293,00	-284,19	3	-228,43	-219,61
Outubro (m=10)			Novembro (m=11)			Dezembro (m=12)		
Ordem	AIC	BIC	Ordem	AIC	BIC	Ordem	AIC	BIC
1	<b>-145,41</b>	<b>-141,00</b>	1*	-77,30	<b>-72,89</b>	1*	-88,65	<b>-93,06</b>
2	-143,42	-136,81	2	<b>-78,00</b>	-71,39	2	-91,41	-84,79
3	-142,41	-133,59	3	-76,05	-67,23	3	<b>-91,82</b>	-83,00

Tabela 5.14: Estimativas pontual e intervalar dos parâmetros do modelo BC-NG-PAR que melhor se ajusta à série de Sobradinho quando utilizamos  $\hat{\lambda} = -0.3$ 

	EMV	IC	EMV	IC	EMV	IC
	Janeiro		Fevereiro		Março	
$\phi_1$	0,5829	(0,5402; 0,6257)	0,4805	(0,2787; 0,6823)	0,7900	(0,7405; 0,8395)
$\sigma_m^*$	0,6354	(0,2336; 1,0372)	1,2882	(0,9160; 1,6605)	0,6537	(0,2377; 1,0697)
s	1,0273	(0,5874; 1,4673)	2,1890	(1,0365; 3,3415)	1,2403	(0,5259; 1,9546)
	Abril		Maio		Junho	
$\phi_1$	0,5957	(0,4097; 0,7817)	0,7634	(0,7346; 0,7922)	0,6964	(0,5373; 0,8555)
$\phi_2$			-0,0997	(-0,1624; -0,0370)	0,1659	(-0,0041; 0,3360)
$\phi_3$			0,2426	(0,2038; 0,2815)	0,1516	(0,0539; 0,2493)
$\sigma_m^*$	1,2317	(1,0056; 1,4579)	0,2605	(0,0344; 0,4866)	0,4537	(0,3331; 0,5744)
s	4,5176	(0,5072; 8,5280)	0,9174	(0,4452; 1,3895)	2,8552	(0,7859; 4,9245)
	Julho		Agosto		Setembro	
$\phi_1$	0,9749	(0,9273; 1,0225)	1,0030	(0,9787; 1,0273)	1,5404	(1,5192; 1,5616)
$\phi_2$					-0,5843	(-0,6058; -0,5629)
$\sigma_m^*$	0,3501	(0,2539; 0,4464)	0,1615	(0,0620; 0,2609)	0,1781	(0,0221; 0,3340)
s	2,8458	(0,6989; 4,9927)	1,0311	(0,5997; 1,4626)	0,8407	(0,4569; 1,2245)
	Outubro		Novembro		Dezembro	
$\phi_1$	0,7823	(0,6800; 0,8845)	0,5782	(0,3667; 0,7897)	0,6100	(0,4375; 0,7826)
$\sigma_m^*$	1,1073	(0,9596; 1,2550)	1,1689	(0,7023; 1,6355)	1,2310	(0,9083; 1,5537)
s	6,9193	(-0,3953; 14,2338)	2,0648	(0,5511; 3,5785)	2,6150	(1,0080; 4,2220)

Notamos que todos os termos autorregressivos são significativos, exceto o segundo termo do mês de junho quando utilizamos a distribuição normal generalizada, mas o limite inferior é muito próximo de zero, por isso o consideramos na modelagem.

Comparando as estimativas pontuais dos termos autorregressivos observamos que as dos meses de fevereiro, julho, agosto, outubro, novembro e dezembro são bem próximas. Verificamos também que a diferença entre as estimativas nos demais meses não é muito grande, comparando os meses que apresentam a mesma quantidade de parâmetros autorregressivos.

Utilizando o critério de seleção AIC para selecionar qual a melhor distribuição para cada mês, verificamos que a distribuição normal generalizada é a mais indicada para os meses janeiro, março, abril, maio, junho, agosto, setembro e outubro. Para os demais meses a distribuição normal é a mais indicada. Adotando o critério BIC observamos



Tabela 5.15: Estimativas pontual e intervalar dos parâmetros do modelo BC-N-PAR que melhor se ajusta à série de Sobradinho quando utilizamos  $\hat{\lambda} = -0.3$ 

	EMV	IC	EMV	IC	EMV	IC
	Janeiro		Fevereiro		Março	
$\phi_1$	0,5018	(0,2948; 0,7088)	0,4834	(0,2731; 0,6936)	0,7412	(0,5796; 0,9028)
$\sigma_m^*$	0,8643	(0,7169; 1,0117)	0,8726	(0,7248; 1,0203)	0,6722	(0,5584; 0,7860)
	Abril		Maio		Junho	
$\phi_1$	0,6913	(0,5238; 0,8589)	0,7631	(0,6117; 0,9145)	0,7904	(0,6313; 0,9496)
$\phi_2$			0,1711	(0,0222; 0,3200)	0,1864	(0,0271; 0,3457)
$\sigma_m^*$	0,6985	(0,5803; 0,8168)	0,4416	(0,3668; 0,5163)	0,3044	(0,2529; 0,3560)
	Julho		Agosto		Setembro	
$\phi_1$	0,9658	(0,9136; 1,0180)	0,9756	(0,9225; 1,0287)	1,4326	(1,0562; 1,8090)
$\phi_2$					-0,5215	(-0,8981; -0,1449)
$\sigma_m^*$	0,2161	(0,1795; 0,2526)	0,2174	(0,1806; 0,2542)	0,3418	(0,2839; 0,3997)
	Outubro		Novembro		Dezembro	
$\phi_1$	0,8069	(0,6566; 0,9572)	0,5819	(0,3866; 0,7771)	0,6262	(0,4393; 0,8131)
$\sigma_m^*$	0,6125	(0,5088; 0,7162)	0,8133	(0,6756; 0,9510)	0,7800	(0,6479; 0,9121)

divergência em relação ao critério AIC nos meses de março e abril, sendo que para o critério BIC a melhor distribuição para estes meses é a distribuição normal.

Consideramos que o ajuste que melhor representa a série de vazões de Sobradinho é adotar a distribuição normal generalizada para os meses de janeiro, maio, agosto, setembro e outubro, e para os demais meses utilizar a distribuição normal. Observamos que, segundo os critérios de seleção, o mês de junho é melhor representado pela distribuição normal generalizada, mas verificamos que o intervalo de confiança para o parâmetro  $s$  contém o valor dois, por isso optamos por adotar a distribuição normal. A ordem para cada mês é mostrado nas Tabelas 5.12 e 5.13.

A partir do modelo com a mescla das distribuições obtemos os resíduos, a partir dele calculamos as funções de autocorrelação periódica e autocorrelação parcial periódica, apresentadas na Figura 5.5, e as funções de autocorrelação e autocorrelação periódica dos resíduos agrupados, mostradas na Figura 5.6.

Notamos por meio das Figuras 5.5 e 5.6 que o ajuste utilizando os modelos BC-NG-PAR e BC-N-PAR eliminou a correlação existente na série de vazões de Sobradinho.

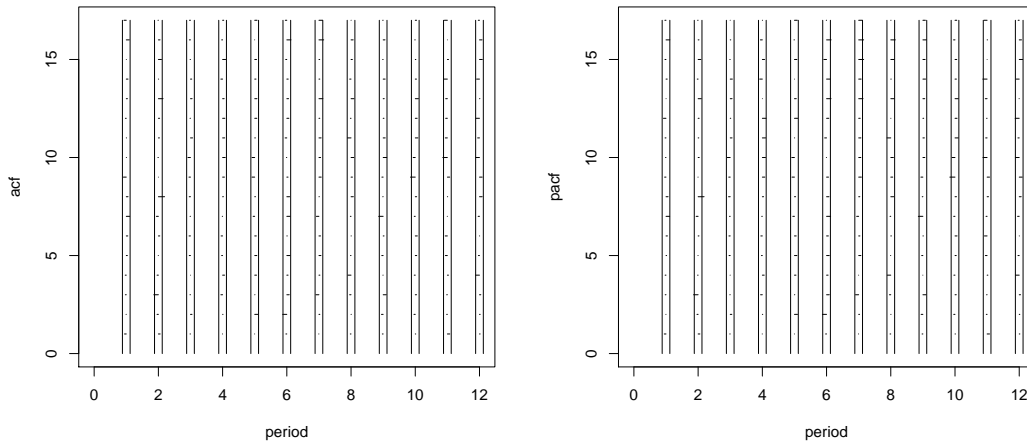


Figura 5.5: Funções de autocorrelação periódica (esquerda) e autocorrelação periódica parcial (direita) dos resíduos do ajuste do modelo com transformação de Box-Cox, as distribuições normal e normal generalizada e  $\hat{\lambda} = -0.3$  à série de vazões de Sobradinho

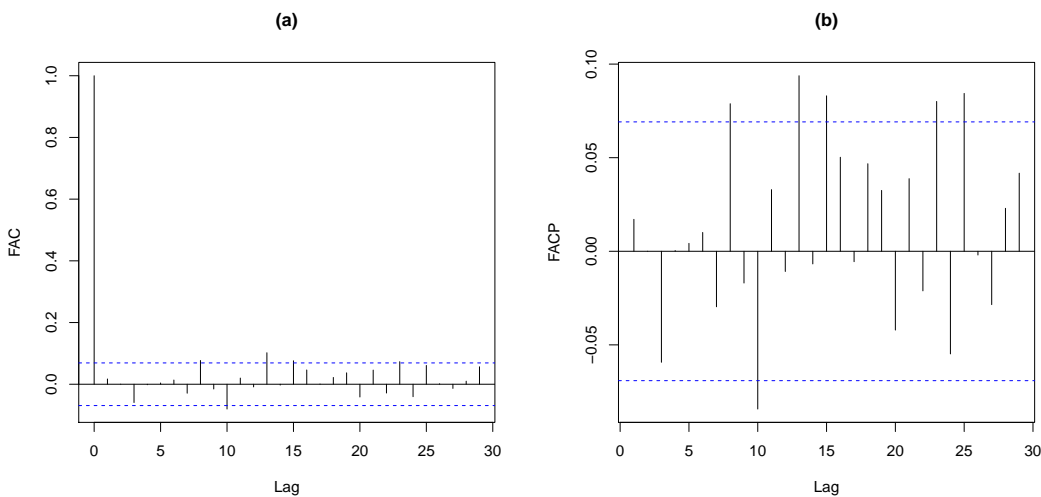


Figura 5.6: Funções de autocorrelação e autocorrelação parcial dos resíduos agrupados do ajuste do modelo com transformação de Box-Cox, as distribuições normal e normal generalizada e  $\hat{\lambda} = -0.3$  à série de vazões de Sobradinho em (a) e em (b), respectivamente

Considerando os dois cenários já adotados anteriormente, calculamos as previsões para o ano de 1998. Na Figura 5.7 apresentamos os resultados.

Comparando as previsões obtidas por meio dos modelos BC-NG-ARMA e BC-NG-PAR, apresentadas nas Figuras 3.14 e 5.7, não notamos diferenças no primeiro cenário. Já no segundo cenário, verificamos diferenças principalmente nos meses de fevereiro,

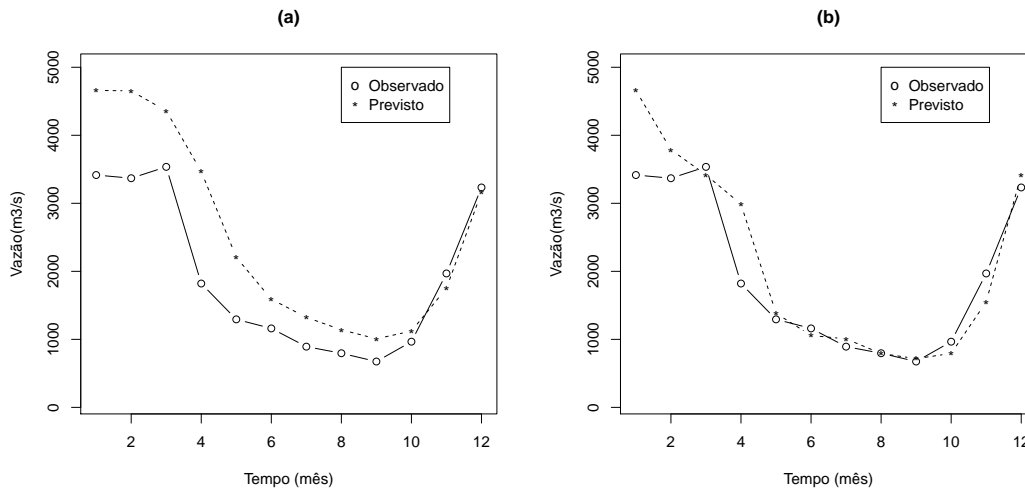


Figura 5.7: Previsões para a série de Sobradinho utilizando o modelo com transformação de Box-Cox, as distribuições normal e normal generalizada e  $\hat{\lambda} = -0.3$  com o primeiro cenário em (a) e o segundo em (b)

março, novembro e dezembro.

Utilizando as métricas já utilizadas neste trabalho, medimos a eficiência das previsões para o ano de 1998. Os resultados estão mostrados na Tabela 5.16. Comparando os resultados destas métricas com os obtidos quando adotamos o modelo BC-NG-ARMA para o ajuste, observamos que em todos os cenários e para todas as métricas o modelo BC-NG-PAR apresentou melhores resultados.

Tabela 5.16: Resumos dos critérios MAPE, EQM e EAM para as previsões da série de Sobradinho utilizando o modelo com transformação de Box-Cox, as distribuições normal e normal generalizada e  $\hat{\lambda} = -0,3$

Cenário	MAPE(%)	EQM	EAM
1º	38,8	674.624,5	657,8
2º	16,4	281.019,1	339,9

Com o interesse em saber o que ocorreria se ao invés de escolher o valor -0,3 para  $\hat{\lambda}$  tivéssemos escolhido o valor -0,1, que são as respectivas estimativas quando utilizamos as distribuições normal generalizada e normal, para repetimos o ajuste considerando  $\hat{\lambda} = -0,1$ .

Comparando as estimativas de máxima verossimilhança dos termos autorregres-

sivos obtidas para os dois valores de  $\hat{\lambda}$  verificamos que, de um modo geral, as estimativas pontuais são próximas, considerando os ajustes com as duas distribuições utilizadas.

No modelo final considerando  $\hat{\lambda} = -0,1$ , a distribuição normal generalizada é indicada para os meses de janeiro, maio, agosto e setembro, e a distribuição normal para os demais meses. Comparando com o outro ajuste observamos que a única alteração é referente ao mês de outubro, pois adotando  $\hat{\lambda} = -0,3$  obtemos que a distribuição normal generalizada é a mais indicada.

As métricas que medem a eficiência das previsões para o ano de 1998 quando adotamos  $\hat{\lambda} = -0,1$  são apresentadas na Tabela 5.17.

Tabela 5.17: Resumos dos critérios MAPE, EQM e EAM para as previsões da série de Sobradinho utilizando o modelo com transformação de Box-Cox, as distribuições normal e normal generalizada e  $\hat{\lambda} = -0,1$

Cenário	MAPE(%)	EQM	EAM
1º	39,8	726.482,9	675,9
2º	16,4	296.510,1	347,2

Comparando as previsões quando adotamos os valores  $-0,3$  e  $-0,1$  para  $\hat{\lambda}$  observamos que os valores preditos são próximos, mas notamos que em todos os cenários e para todas as métricas o modelo final utilizando  $\hat{\lambda} = -0,3$  apresenta resultados melhores, exceto no segundo cenário e para a métrica MAPE, neste caso os valores das métricas são iguais.

## 5.3 Conclusão

Neste capítulo apresentamos um novo modelo, chamado de BC-NG-PAR, para séries temporais que apresentam periodicidade na média, na variância e na função de autocorrelação, sendo definido também um caso particular, denominado de LNG-PAR. Para o modelo geral, encontramos as estimativas de máxima verossimilhança utilizando dois passos, analisamos este procedimento através de um estudo de simulação que apresentou ótimos resultados. Mostramos também um método para calcular as previsões a partir do modelo proposto.

---

Exemplificamos duas aplicações dos modelos sugeridos utilizando as séries de vazões das usinas hidrelétricas de Furnas e de Sobradinho, já estudadas neste trabalho. As previsões calculadas a partir dos modelos apresentados neste capítulo mostraram melhores resultados comparado com as previsões obtidas por meio dos modelos mostrados nos capítulos anteriores.

# Capítulo 6

## Modelo Normal Generalizada Zero Inflacionada ARMA

São vários os conjuntos de dados que apresentam excesso de zero, neste capítulo estamos interessados em analisar uma série temporal resultante da observação diária da quantidade de precipitação pluviométrica. A série em questão apresenta uma grande quantidade de observações cujo valor é zero, além de apresentar correlação entre os valores, dificultando assim a análise da mesma.

Mostramos neste capítulo a definição da distribuição normal generalizada zero inflacionada na Seção 6.1, juntamente com a função de máxima verossimilhança. Desenvolvemos, a partir da distribuição proposta, um modelo que pode ser utilizado para modelar séries temporais que apresentam excesso de zero. Este é apresentado na Seção 6.2. Nesta mesma Seção apresentamos a função de verossimilhança do modelo proposto. Exemplificamos uma aplicação do modelo a partir da série da quantidade de precipitação pluviométrica diária da cidade de São Carlos, que é mostrado na Seção 6.2.2. Algumas discussões e conclusões são expostas na Seção 6.3.

## 6.1 Distribuição Normal Generalizada Zero Inflacionada

Uma variável aleatória  $X$  é dita ter distribuição normal generalizada zero inflacionada (NGZI) se sua função densidade de probabilidade é definida por

$$f_{NGZI}(x|p, \mu, \sigma, s) = p^{I_{\{0\}}(x)}(1-p)^{1-I_{\{0\}}(x)}\{f_{NG}(x|\mu, \sigma, s)\}^{1-I_{\{0\}}(x)}, \quad (6.1)$$

sendo que o parâmetro  $p$  é restrito por  $0 \leq p \leq 1$ ,  $f_{NG}(x|\mu, \sigma, s)$  é a função densidade de probabilidade da distribuição normal generalizada com os parâmetros  $\mu$ ,  $\sigma$  e  $s$  ( $NG(\mu, \sigma, s)$ ) e  $I_{\{0\}}(x)$  é uma função indicadora, definida como

$$I_{\{0\}}(x) = \begin{cases} 1 & \text{se } x = 0 \\ 0 & \text{se } x \neq 0. \end{cases} \quad (6.2)$$

Interpretamos a função densidade de probabilidade  $f_{NGZI}(\cdot)$  como sendo uma mistura da densidade de probabilidade da variável aleatória normal generalizada com uma função com toda massa de probabilidade no ponto zero. Desta forma,  $f_{NGZI}(0) = p$  e  $f_{NGZI}(x) = (1-p)f_{NG}(x)$ , para todo  $x \in \mathbb{R} - \{0\}$ . Pode ser mostrado por integração que  $f_{NGZI}(\cdot)$  é uma função densidade de probabilidade, ou seja, tem integral igual 1 no suporte  $\Omega_X = \{x \in \mathbb{R}\}$ .

A distribuição  $f_{NGZI}(\cdot)$  pode ser útil na análise de conjuntos de dados reais cujas observações seguem uma distribuição normal generalizada e apresente uma grande quantidade de observações com valor zero.

### 6.1.1 Função de Verossimilhança

Seja  $\mathbf{x} = \{x_1, \dots, x_n\}$  uma amostra aleatória da distribuição normal generalizada zero inflacionada. A função de verossimilhança associada a  $\mathbf{x}$  para obter as estimativas de máxima verossimilhança dos parâmetros de interesse  $p$ ,  $\mu$ ,  $\sigma$  e  $s$  é dada por

$$L(p, \mu, \sigma, s|\mathbf{x}) = \prod_{i=1}^n p^{\epsilon_i}(1-p)^{1-\epsilon_i}\{f_{NG}(x_i|\mu, \sigma, s)\}^{1-\epsilon_i}, \quad (6.3)$$

em que  $\epsilon_i = I_{\{0\}}(x_i)$ , ou seja,  $\epsilon_i = 1$  se  $x_i = 0$  e  $\epsilon_i = 0$  se  $x_i \neq 0$ . A função log-verossimilhança é dada por

$$\begin{aligned} l(p, \mu, \sigma, s|\mathbf{x}) &= \sum_{i=1}^n \epsilon_i \log(p) + \sum_{i=1}^n (1 - \epsilon_i) \log(1 - p) + \sum_{i=1}^n (1 - \epsilon_i) \log f_{NG}(x_i|\mu, \sigma, s) \\ &= \sum_{i=1}^n \epsilon_i \log\left(\frac{p}{1-p}\right) + n \log(1 - p) + \sum_{i=1}^n (1 - \epsilon_i) \log f_{NG}(x_i|\mu, \sigma, s). \end{aligned} \quad (6.4)$$

Notemos que a função log-verossimilhança pode ser escrita como

$$l(p, \mu, \sigma, s|\mathbf{x}) = l(p|\mathbf{x}) + l(\mu, \sigma, s|\mathbf{x}^+),$$

em que  $\mathbf{x}^+$  é o vetor contendo somente as observações positivas do vetor de observações  $\mathbf{x}$ . As funções  $l(p|\mathbf{x})$  e  $l(\mu, \sigma, s|\mathbf{x}^+)$  são dadas respectivamente por

$$l(p|\mathbf{x}) = \sum_{i=1}^n \epsilon_i \log\left(\frac{p}{1-p}\right) + n \log(1 - p), \quad (6.5)$$

$$l(\mu, \sigma, s|\mathbf{x}^+) = \sum_{x_i > 0} \log f_{NG}(x_i), \quad (6.6)$$

portanto as funções  $l(p|\mathbf{x})$  e  $l(\mu, \sigma, s|\mathbf{x}^+)$  podem ser utilizadas para a obtenção das estimativas de máxima verossimilhança do parâmetro  $p$  separadamente dos parâmetros  $\mu$ ,  $\sigma$  e  $s$ .

O estimador de  $p$  pode ser encontrado de maneira fechada, pois observamos que

$$\frac{dl(p|\mathbf{x})}{dp} = \frac{n_0}{p} + \frac{n_0}{1-p} - \frac{n}{1-p} \quad (6.7)$$

sendo que  $n_0$  é a quantidade de vezes que foram observados valores zero na amostra. Igualando a zero a equação (6.7) obtemos um candidato a estimador de máxima verossimilhança dado por

$$\hat{p} = \frac{n_0}{n}. \quad (6.8)$$

A segunda derivada em relação a  $p$  é dada por

$$\frac{d^2l(p|\mathbf{x})}{dp^2} = -\frac{n_0}{p^2} + \frac{n_0}{(1-p)^2} - \frac{n}{(1-p)^2}, \quad (6.9)$$

confirmamos que o candidato é mesmo o estimador de máxima verossimilhança, pois a segunda derivada é negativa para todo  $0 < p < 1$ .



Observamos que a função de verossimilhança para  $(\mu, \sigma, s)$  é bastante complexa, por isso indicamos métodos iterativos para a obtenção das estimativas de máxima verossimilhança.

## 6.2 Modelo Normal Generalizada Zero Inflacionada ARMA

Considere  $\{X_t; t \in \mathbb{N}\}$  um processo estocástico, suponhamos que a distribuição condicional do processo dada toda a informação passada possui função densidade de probabilidade normal generalizada zero inflacionada, ou seja,

$$f_{X_t|\mathfrak{F}_{t-1}}(x_t|p_t, \mu_t, \sigma, s) = p_t^{I_{\{0\}}(x_t)}(1 - p_t)^{1 - I_{\{0\}}(x_t)} \{f_{NG}(x_t|\mu_t, \sigma, s)\}^{1 - I_{\{0\}}(x_t)}, \quad (6.10)$$

sendo que  $\mathfrak{F}_{t-1}$  é o conjunto de informações das observações até o instante  $t - 1$ , definido como

$$\mathfrak{F}_{t-1} = \{x_1, \dots, x_{t-1}, p_1, \dots, p_{t-1}, \mu_1, \dots, \mu_{t-1}\}. \quad (6.11)$$

Observamos que os parâmetros  $p_t$  e  $\mu_t$  variam com o tempo. Como o interesse é inserir uma estrutura de correlação nesses parâmetros, assumimos que  $p_t$  se relaciona com os termos autorregressivos e de médias móveis por uma função de ligação,  $g$ , monótona, bijetiva e duas vezes diferenciável, sendo dada por

$$g(p_t) = \alpha_0 + \sum_{i=1}^{p^*} \alpha_i a(x_{t-i}) - \sum_{j=1}^{q^*} \gamma_j b(p_{t-j}) \quad (6.12)$$

sendo  $g$ ,  $a$  e  $b$  funções apropriadas. Adotamos para as funções  $g$  e  $b$  a função logito e para a função  $a$  a composição da função logarítmica com a inversa da função logito que resulta em

$$a(x_t) = \log \left( \frac{e^{x_t}}{1 + e^{x_t}} \right),$$

logo, a equação (6.12) fica dada por

$$g(p_t) = \alpha_0 + \sum_{i=1}^{p^*} \alpha_i \log \left( \frac{e^{x_{t-i}}}{1 + e^{x_{t-i}}} \right) - \sum_{j=1}^{q^*} \gamma_j \log \left( \frac{p_{t-j}}{1 - p_{t-j}} \right). \quad (6.13)$$

Assumimos que  $\mu_t$  é dado da mesma maneira como no Capítulo 2, quando não admitimos variáveis exógenas, ou seja,

$$\mu_t = \phi_0 + \sum_{i=1}^p \phi_i x_{t-i} + \sum_{j=1}^q \theta_j (x_{t-j} - \mu_{t-j}). \quad (6.14)$$

### 6.2.1 Função de Verossimilhança

Denotemos por  $\mathbf{x} = \{x_1, \dots, x_n\}$  a série temporal observada. Assumindo que os dados observados seguem o modelo normal generalizada zero inflacionada ARMA, os parâmetros a serem estimados são  $(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s)$  sendo que  $\boldsymbol{\alpha} = \alpha_0, \dots, \alpha_{p^*}$ ,  $\boldsymbol{\gamma} = \gamma_1, \dots, \gamma_{q^*}$ ,  $\boldsymbol{\phi} = \phi_0, \dots, \phi_p$  e  $\boldsymbol{\theta} = \theta_1, \dots, \theta_q$ . Sejam  $r^* = \max(p^*, q^*, p, q)$  e  $\mathbf{x}_{r^*}$  as  $r^*$ -primeiras observações, ou seja,  $\mathbf{x}_{r^*} = \{x_1, \dots, x_{r^*}\}$ , a função de verossimilhança condicionada nas  $r^*$ -primeiras observações é dada por

$$L(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \mathbf{x}) = \prod_{t=r^*+1}^n p_t^{\epsilon_t} (1-p_t)^{1-\epsilon_t} \{f_{NG}(x_t | \mu_t, \sigma, s)\}^{1-\epsilon_t}, \quad (6.15)$$

em que  $\epsilon_t = I_{\{0\}}(x_t)$ . A função log-verossimilhança é dada por

$$\begin{aligned} l(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \mathbf{x}) &= \sum_{t=r^*+1}^n \epsilon_t \log(p_t) + \sum_{t=r^*+1}^n (1-\epsilon_t) \log(1-p_t) \\ &\quad + \sum_{t=r^*+1}^n (1-\epsilon_t) \log f_{NG}(x_t | \mu_t, \sigma, s). \end{aligned} \quad (6.16)$$

Notemos que de maneira análoga à função log-verossimilhança descrita na equação (6.4), podemos reescrever a equação (6.16) como

$$l(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \mathbf{x}) = l(\boldsymbol{\alpha}, \boldsymbol{\gamma} | \mathbf{x}) + l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \mathbf{x}^+),$$

em que  $\mathbf{x}^+$  como definido anteriormente. As funções  $l(\boldsymbol{\alpha}, \boldsymbol{\gamma} | \mathbf{x})$  e  $l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \mathbf{x}^+)$  são dadas por

$$l(\boldsymbol{\alpha}, \boldsymbol{\gamma} | \mathbf{x}) = \sum_{t=r^*+1}^n \epsilon_t \log \left( \frac{p_t}{1-p_t} \right) + \sum_{t=r^*+1}^n \log(1-p_t), \quad (6.17)$$

$$l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \mathbf{x}^+) = \sum_{x_t > 0} \log f_{NG}(x_t), \quad (6.18)$$

em que  $p_t$  e  $\mu_t$  como descrito nas equações (6.13) e (6.14), respectivamente. Mesmo dividindo a função log-verossimilhança em duas funções, estas continuam apresentando alta complexidade. Consequentemente, utilizamos métodos iterativos para a obtenção das estimativas de máxima verossimilhança. Construimos os intervalos de confiança adotando normalidade assintótica dos estimadores de máxima verossimilhança.

## 6.2.2 Aplicação

Consideramos a série temporal em que foi observado a quantidade de precipitação pluviométrica diária da cidade de São Carlos - São Paulo - Brasil, entre os dias 01/01/2000 e 10/10/2013. A série temporal foi obtida através do site do Instituto Nacional de Meteorologia do Brasil (<http://www.inmet.gov.br>), e é apresentada na Figura 6.1.

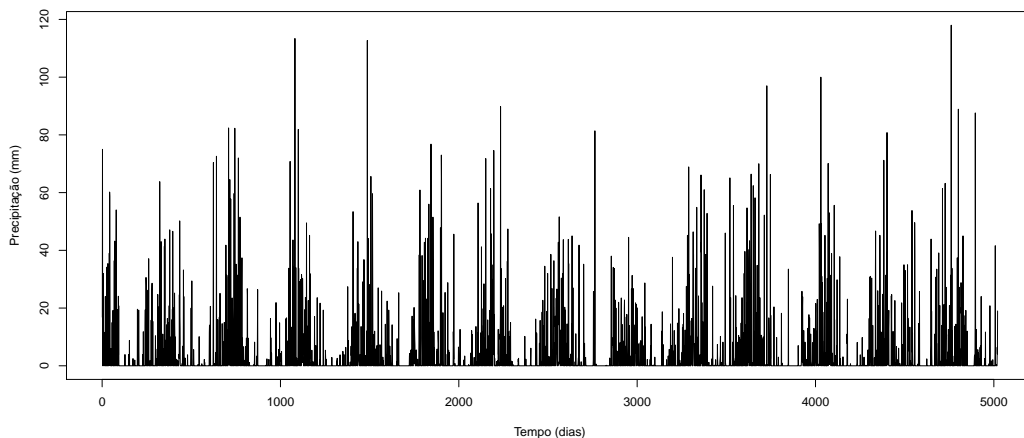


Figura 6.1: Quantidade de precipitação pluviométrica diária da cidade de São Carlos

Analisamos a correlação existente na série por meio dos gráficos das funções de autocorrelação e autocorrelação parcial, que são mostrados na Figura 6.2. Por ela observamos que a série apresenta correlação.

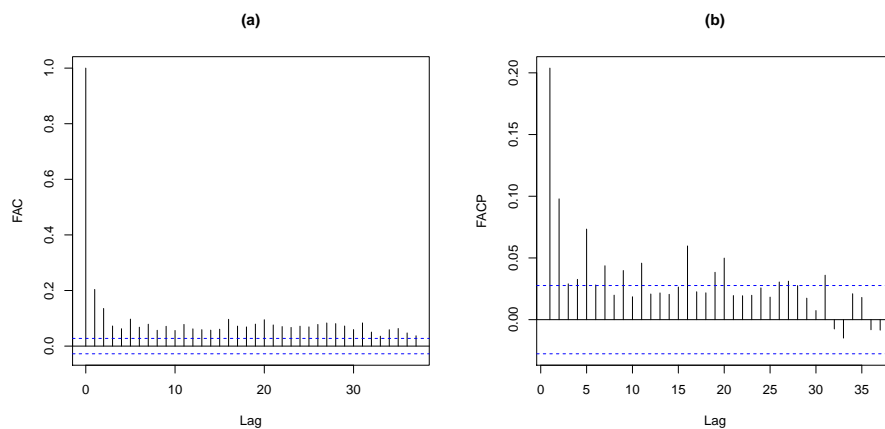


Figura 6.2: Funções de autocorrelação e autocorrelação parcial da série de precipitação pluviométrica diária da cidade de São Carlos em (a) e (b), respectivamente

Adotando o modelo normal generalizada zero inflacionada ARMA para modelar a série da quantidade de precipitação pluviométrica temos que  $p_t$  é a probabilidade de não ocorrer chuva, e quando chover  $\mu_t$  representa a quantidade de precipitação.

Como apresentado na Seção 6.2.1 podemos ajustar o modelo em duas partes. Para a modelagem de  $\mu_t$  criamos uma série auxiliar com apenas os valores maiores que zero da série original, para um melhor ajuste utilizamos o logaritmo da série auxiliar, analisamos a correlação desta série na Figura 6.3.

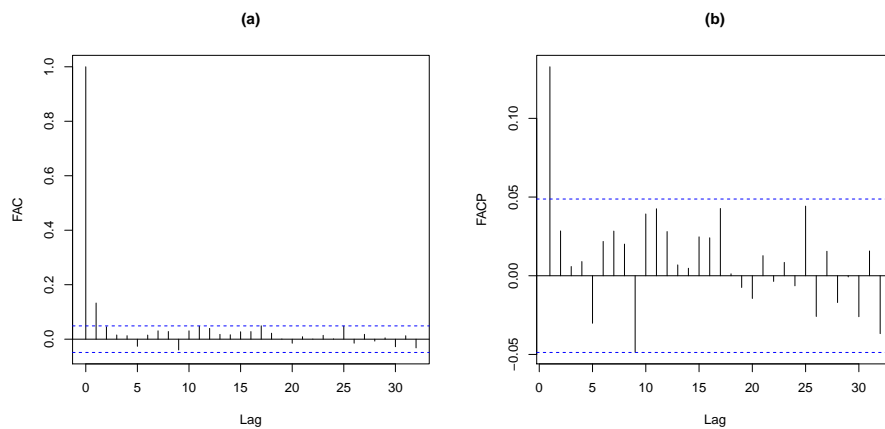


Figura 6.3: Funções de autocorrelação e autocorrelação parcial do logaritmo da série de precipitação pluviométrica diária positiva da cidade de São Carlos em (a) e (b), respectivamente

Adotamos os critérios AIC e BIC para a escolha da melhor ordem, utilizamos os modelos NGZI-ARMA e também o caso particular NZI-ARMA, os resultados dos critérios são apresentados na Tabela 6.1.

Por meio da Tabela 6.1 observamos que ambos os critérios indicaram que o melhor ajuste é utilizando a distribuição normal generalizada e com um termo autorregressivo.

As estimativas de máxima verossimilhança e os intervalos de confiança assumindo normalidade assintótica do modelo indicado, por meio dos critérios, são mostrados na Tabela 6.2.

Notemos que a estimativa pontual do parâmetro  $s$  é maior do que 2, ou seja, o ajuste indicou que a série temporal apresenta caudas mais leves do que as da distribuição normal. Observamos também que seu intervalo de confiança assintótico não contém o valor dois, indicando assim que a distribuição normal generalizada apresenta melhores

Tabela 6.1: Seleção entre os modelos NGZI-ARMA e NZI-ARMA utilizando os critérios AIC e BIC para explicar a quantidade de precipitação  $\mu_t$ 

Modelo	AIC	BIC
NGZI-ARMA(1,0)	<b>5885,72</b>	<b>5907,27</b>
NGZI-ARMA(2,0)	5891,70	5918,65
NGZI-ARMA(0,1)	5886,70	5908,26
NGZI-ARMA(1,1)	5887,27	5914,21
NZI-ARMA(1,0)	5915,26	5931,43
NZI-ARMA(0,1)	5916,76	5932,92
NZI-ARMA(1,1)	5916,61	5938,16

Tabela 6.2: Estimativas pontual e intervalar dos parâmetros do modelo NGZI-ARMA(1,0) para explicar a quantidade de precipitação  $\mu_t$ 

Parâmetro	EMV	IC
$\phi_0$	1,3513	(1,2417; 1,4610)
$\phi_1$	0,1141	(0,0686; 0,1597)
$\sigma$	2,4811	(2,3519; 2,6103)
$s$	3,0357	(2,5671; 3,5044)

resultados do que a distribuição normal.

Analisamos o ajuste da distribuição normal generalizada ao logaritmo da série positiva e maior que zero através dos resíduos. Usamos as funções de autocorrelação e autocorrelação parcial, estas estão apresentadas na Figura 6.4. Por meio dos gráficos observamos que o ajuste eliminou a correlação existente na série.

Investigamos a probabilidade de não ocorrer chuva,  $p_t$ , através das funções de autopersistência, introduzidas por Startz (2012), que são dadas por

$$APF^0(k) = P(y_{t+k} = 1 | y_t = 0), \quad (6.19)$$

$$APF^1(k) = P(y_{t+k} = 1 | y_t = 1), \quad (6.20)$$

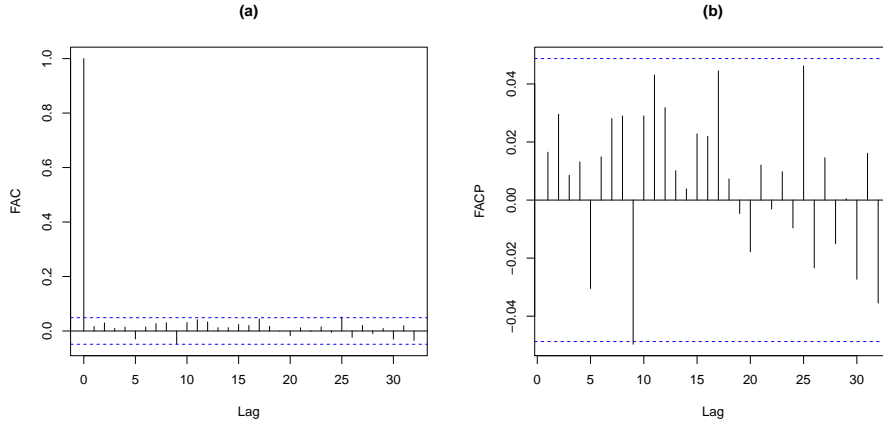


Figura 6.4: Funções de autocorrelação e autocorrelação parcial dos resíduos do ajuste do modelo NGZI-ARMA(1,0) para explicar a quantidade de precipitação  $\mu_t$  em (a) e (b), respectivamente

que é estimada, segundo Wang e Li (2011), por

$$APF^0(k) = \frac{n}{n-k} \frac{\sum_{t=1}^{n-k} I\{y_{t+k} = 1, y_t = 0\}}{\sum_{t=1}^n I\{y_t = 0\}}, \quad (6.21)$$

$$APF^1(k) = \frac{n}{n-k} \frac{\sum_{t=1}^{n-k} I\{y_{t+k} = 1, y_t = 1\}}{\sum_{t=1}^n I\{y_t = 1\}}, \quad (6.22)$$

sendo que  $I\{y_{t+k} = i, y_t = j\}$  e  $I\{y_t = i\}$  são funções indicadoras. Para analisar  $p_t$  criamos uma nova série auxiliar, essa série recebe o valor zero quando não houve precipitação pluviométrica, e recebe o valor um quando ocorreu precipitação, portanto, a série auxiliar é uma série binária. O gráfico com as funções de autopersistência é apresentado na Figura 6.5.

Observamos por meio da Figura 6.5 que a probabilidade de não ocorrer chuva apresenta um comportamento sazonal, por isso substituímos o parâmetro  $\alpha_0$  da equação (6.13) por duas funções periódicas, resultando em

$$g(p_t) = \alpha_1 \cos(2\pi f_t) + \alpha_2 \sin(2\pi f_t) + \sum_{i=1}^{p^*} \alpha_i \log \left( \frac{e^{x_{t-i}}}{1 + e^{x_{t-i}}} \right) - \sum_{j=1}^{q^*} \gamma_j \log \left( \frac{p_{t-j}}{1 - p_{t-j}} \right),$$

sendo  $f_t = \frac{t}{365}$ , assim esperamos capturar a sazonalidade da não probabilidade de chuva.

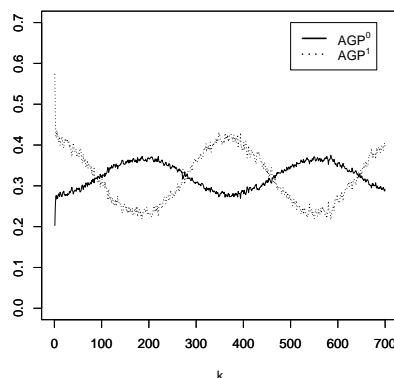


Figura 6.5: Funções de autopersistência estimada para a série que indica se ocorreu chuva ou não

Novamente adotamos os critérios AIC e BIC para a escolha da melhor ordem, apresentamos os resultados dos critérios para a escolha do melhor modelo para explicar  $p_t$  na Tabela 6.3.

Tabela 6.3: Seleção da melhor ordem do modelo NGZI-ARMA utilizando os critérios AIC e BIC para explicar a probabilidade de não ocorrer chuva  $p_t$

Modelo	AIC	BIC
NGZI-ARMA(1,0)	5323,35	<b>5342,91</b>
NGZI-ARMA(0,1)	6327,98	6347,55
NGZI-ARMA(1,1)	5322,95	5349,04
NGZI-ARMA(2,0)	<b>5322,18</b>	5348,27

Por meio da Tabela 6.3 observamos que o critério AIC indicou que o melhor ajuste é o modelo de ordem (2,0), já o critério BIC apontou o modelo de ordem (1,0). Optamos por adotar o modelo escolhido por meio do critério BIC como sendo o melhor modelo que se ajusta à série, pois o modelo com ordem (2,0) apresenta o segundo termo autorregressivo não significativo. Apresentamos as estimativas de máxima verossimilhança e os intervalos de confiança, assumindo normalidade assintótica, na Tabela 6.4.

Notamos que as estimativas pontuais apresentadas na Tabela 6.4 são todas negativas, a partir das estimativas intervalares observamos que todas as estimativas são significativas.

Tabela 6.4: Estimativas pontuais e intervalar dos parâmetros do modelo NGZI-ARMA(1,0) que melhor explicou a probabilidade de não ocorrer chuva  $p_t$

Parâmetro	EMV	IC
$\alpha_1$	-0,9755	(-1,0714; -0,8796)
$\alpha_2$	-0,1992	(-0,2905; -0,1080)
$\phi_1$	-1,7860	(-1,9062; -1,6658)

Interpretando os resultados, observamos que a probabilidade de não ocorrência de chuva em um determinado dia é correlacionada com a precipitação do dia anterior e qual o dia do ano. A quantidade de precipitação de um dia chuvoso é relacionado com a quantidade de precipitação do último dia chuvoso.

A probabilidade de não chover do ano de 2000 estimada por meio do modelo NGZI-ARMA(1,0) é apresentada na Figura 6.6. Observamos que a probabilidade de não chover é menor no começo e no final do ano e que a ocorrência de chuva em um determinado dia faz com que a probabilidade de não ocorrer precipitação no dia seguinte diminua, e esta diminuição depende da quantidade pluviométrica do dia anterior.

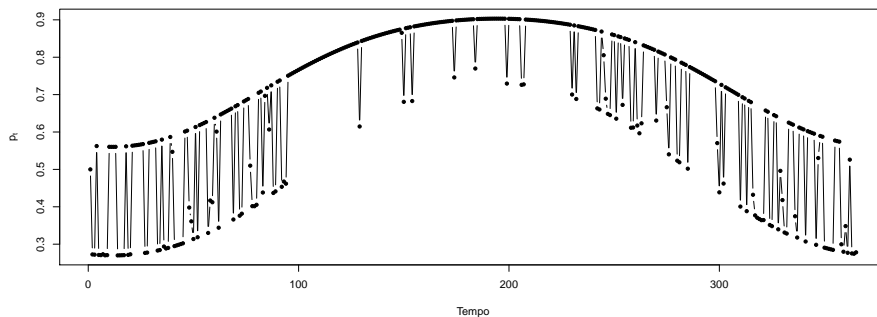


Figura 6.6: Probabilidade de não ocorrer chuva do ano de 2000 estimada através do modelo NGZI-ARMA(1,0)

## 6.3 Conclusão

A distribuição normal generalizada zero inflacionada e o modelo para séries temporais normal generalizada zero inflacionada ARMA são as propostas apresentadas neste capítulo. Mostramos também as funções de verossimilhança, sendo que em ambos os



---

casos conseguimos dividir a função log-verossimilhança em duas funções, a partir de cada uma delas obtemos as estimativas de máxima verossimilhança para um subconjunto dos parâmetros a serem estimados.

A partir da série temporal em que foi observada a quantidade de precipitação pluviométrica diária da cidade de São Carlos - São Paulo - Brasil, ajustamos o modelo proposto e observamos que a probabilidade de não ocorrer chuva em um determinado dia é altamente correlacionada com a precipitação pluviométrica do dia anterior.

# Capítulo 7

## Conclusões e Propostas Futuras

A opção pela escolha da distribuição normal generalizada é devida à possibilidade das suas caudas serem mais leves ou mais pesadas do que as da distribuição normal. Na maioria das aplicações desenvolvidas aqui, as séries apresentam caudas mais pesadas que as da distribuição normal, sendo assim outras distribuições poderiam ser utilizadas, como por exemplo a distribuição t de Student. Mas observamos que a distribuição da quantidade de precipitação pluviométrica dos dias chuvosos apresenta caudas mais leves comparadas com as da distribuição normal, com isso a formulação do modelo, por exemplo, com distribuição t de Student não é adequado. Por isso optamos pela distribuição normal generalizada, assim os modelos propostos podem ser adotados independentemente do comportamento das caudas da distribuição da série.

No estudo de simulação do modelo NG-ARMA observamos a influência do valor do parâmetro  $s$  sobre as métricas EQM e VAM dos demais parâmetros. Esse comportamento também foi notado no estudo de simulação dos modelos BC-NG-ARMA e LNG-ARMA, mas apenas para o parâmetro  $\sigma$ . Para os modelos NG-ARMA, LNG-ARMA e BC-NG-ARMA os ajustes às séries temporais reais apresentaram melhores resultados comparados ao tradicional modelo ARMA com distribuição normal.

A extensão do modelo periódico autorregressivo considerando a distribuição normal generalizada, chamado de NG-PAR, mostrou a partir da aplicação à série de vazões da usina hidrelétrica de Furnas que seis meses apresentaram caudas mais pesadas do que as da distribuição normal. Quando consideramos o modelo mais robusto, ou seja, adotando

a transformação da série a partir do modelo BC-NG-PAR, notamos a diminuição de meses com curtose diferente da distribuição normal, mas não a total eliminação, indicando que mesmo com a transformação da série existe a necessidade da utilização da distribuição mais flexível.

Adotando as métricas MAPE, EQM e EAM para a comparação das previsões obtidas para as séries de vazões das usinas hidrelétricas de Furnas e Sobradinho por meio dos modelos NG-ARMA, NG-PAR, BC-NG-ARMA e BC-NG-PAR, notamos que as previsões a partir dos modelos periódicos apresentaram menor erro nas métricas adotadas. Portanto, a partir dessas duas séries observamos que uma melhor modelagem da estrutura de autocorrelação produziu melhores previsões.

Analisamos a série da quantidade de precipitação pluviométrica da cidade de São Carlos adotando o modelo proposto NGZI-ARMA e obtivemos que a distribuição da quantidade de precipitação pluviométrica dos dias chuvosos apresenta curtose diferente da distribuição normal e a probabilidade de não ocorrer chuva é correlacionado com a precipitação pluviométrica do dia anterior.

Algumas das propostas futuras de pesquisas são:

1. Apresentar uma abordagem bayesiana para o modelo NG-ARMA, adotando por exemplo as distribuições *a priori* dadas por  $\pi(\beta_0), \pi(\phi_1), \dots, \pi(\phi_p), \pi(\theta_1), \dots, \pi(\theta_q) \propto$  constante,  $\pi(\sigma) \propto \frac{1}{\sigma}$ ,  $\pi(s) \propto \frac{1}{s}$ , com isso a distribuição *a priori* conjunta é dada por  $\pi(\beta_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s) \propto \frac{1}{s\sigma}$ .

A abordagem bayesiana pode trazer como benefício a eliminação da influência do parâmetro  $s$  nas métricas vício absoluto médio e erro quadrático médio dos demais parâmetros, além da utilização da distribuição *a posteriori* preditiva para as previsões, considerando estimativas pontuais e intervalares.

2. Propor uma abordagem bayesiana para o modelo BC-NG-ARMA, utilizando as mesmas distribuições *a priori* para o vetor  $(\beta_0, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma, s)$  e adotando para  $\lambda$  a distribuição *a priori*  $U(-1, 1)$ .

Neste cenário a abordagem bayesiana elimina a imposição da discretização do parâmetro  $\lambda$  e da necessidade da estimação em dois passos, podendo eliminar a influência do parâmetro  $s$  nas métricas vício absoluto médio e erro quadrático médio do parâ-

metro  $\sigma$ , além dos benefícios relacionados à previsão já citados no item anterior.

3. Investigar a eficiência dos métodos bootstrap para séries temporais quando adotados em séries periódicas. Utilizar o método bootstrap para melhorar as estimativas pontuais e intervalares dos modelos NG-PAR e BC-NG-PAR. Também usar o método bootstrap para apresentar intervalos de confiança para as previsões.
4. Desenvolver um método para calcular as previsões a partir do modelo NGZI-ARMA e produzir um estudo de simulação para avaliar as estimativas de máxima verossimilhança utilizando, por exemplo, as métricas probabilidade de cobertura, vício absoluto médio e erro quadrático médio.

# Referências Bibliográficas

- BAUWENS, L.; HAUTSCH, N. Stochastic conditional intensity processes. *Journal of Financial Econometrics*, v. 4(3), p. 450–493, 2006.
- BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, D. M. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, v. 98, p. 214–223, 2003.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, v. 31, p. 307–327, 1986.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society*, v. 26, p. 211–243, 1964.
- BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. Holden-Day, São Francisco, 1976.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. *Time series analysis: forecasting and control*. John Wiley & Sons, New Jersey, 2008.
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to Time Series and Forecasting*. Springer, New York, 2 ed., 2002.
- BYRD, R. H.; LU, P.; NOCEDAL, J.; ZHU, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing*, v. 16, p. 1190–1208, 1995.
- CAEIRO, F.; MATEUS, A. *randtests: Testing randomness in R*, 2014. URL <http://CRAN.R-project.org/package=randtests>. R package version 1.0.
- CAINES, P. E. *Linear Stochastic Systems*. John Wiley & Sons, New York, 1988.

- CORDEIRO, G. M.; ANDRADE, M. G. Transformed generalized linear models. *Journal of Statistical Planning and Inference*, v. 139, p. 2970–2987, 2009.
- COX, D. R. Partial likelihood. *Biometrika*, v. 62, p. 69–76, 1975.
- COX, D. R. Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics, Series B*, v. 20(2), p. 215–242, 1981.
- COX, D. R.; STUART, A. Some quick sign test for trend in location and dispersion. *Biometrika*, v. 42, p. 80–95, 1955.
- DAVIS, R. A.; WU, R. A negative binomial model for time series of counts. *Biometrika*, v. 96, p. 735–749, 2009.
- DAVIS, R. A.; DUNSMUIR, W. T. M.; WANG, Y. On autocorrelation in a Poisson regression model. *Biometrika Trust*, v. 87, p. 491–505, 2000.
- DAVIS, R. A.; DUNSMUIR, W. T. M.; STREETT, S. B. Observation-driven models for poisson counts. *Biometrika*, v. 90, p. 777–790, 2003.
- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, v. 50(4), p. 987–1007, 1982.
- FLORES, B. E. A pragmatic view of accuracy measurement in forecasting. *Omega (Oxford)*, v. 14(2), p. 93–98, 1986.
- GONÇALVES, S.; MEDDAHI, N. Box-Cox transforms for realized volatility. *Journal of Econometrics*, v. 160, p. 129–144, 2011.
- GUPTA, R. C.; LVIN, S. Reliability functions of generalized log-normal model. *Mathematical and Computer Modelling*, v. 42, p. 939–946, 2005.
- HARVEY, A. C. *Time Series Models*. MIT Press, New York, 2 ed., 1993.
- HIPEL, K. W.; MCLEOD, A. I. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, Amsterdam, 1994.
- HURD, H. L.; GERR, N. L. Graphical methods for determining the presence of periodic correlation. *Journal of Time Series Analysis*, v. 12, p. 337–350, 1991.

- JOSE, K. K.; TOMY, L.; SREEKUMAR, J. Autoregressive processes with normal-Laplace marginals. *Statistics and Probability Letters*, v. 78, p. 2456–2462, 2008.
- MARTÍN, J.; PÉREZ, C. J. Bayesian analysis of a generalized lognormal distribution. *Computational Statistics and Data Analysis*, v. 53, p. 1377–1387, 2009.
- MCCULLAGH, P.; NELDER, J. A. *Generalized Linear Models*. Chapman and Hall, London, 1989.
- MCLEOD, A. I. Parsimony, model adequacy and periodic correlation in time series forecasting. *Biometrika*, v. 61, p. 387–393, 1993.
- MILANI, E. A.; ANDRADE, M. G.; DINIZ, C. A. R. Generalized normal arma model applied to the areas of economy, hydrology and public policy. *Communications in Statistics - Simulation and Computation*, 2015.
- MONDAL, M. S.; WASIMI, S. A. Generating and forecasting monthly flows of the Ganges river with par model. *Journal of Hydrology*, v. 323, p. 41–56, 2006.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the Theory of Statistics*. McGraw-Hill, Tokyo, 1974.
- MORETTIN, P. A. *Econometria Financeira*. Edgard Blucher, São Paulo, 2008.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. Edgard Blucher, São Paulo, 2006.
- NADARAJAH, S. A generalized normal distribution. *Journal of Applied Statistics*, v. 32, p. 685–694, 2005.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society*, v. 135, p. 370–384, 1972.
- OLIVEIRA, F. L. C.; SOUZA, R. C. A new approach to identify the structural order of PAR(p) models. *Pesquisa Operacional*, v. 31, p. 487–498, 2011.
- PAGANO, M. On periodic and multiple autoregressions. *The Annals of Statistics*, v. 6, p. 1310–1317, 1978.

- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL <http://www.R-project.org/>.
- ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. *Test*, v. 18, p. 529–545, 2009.
- SHAO, J. *Mathematical Statistics*. Springer, New York, 2 ed., 2003.
- SHEPHARD, N. *Stochastic Volatility. Selected Readings*. Oxford University Press, Oxford, 2005.
- SHUMWAY, R. H.; STOFFER, D. S. *Time Series Analysis and Its Applications: With R Examples*. Springer, New York, 2011.
- SINGH, B.; SHARMA, K. K.; RATHI, S.; SINGH, G. A generalized log-normal distribution and its goodness of fit to censored data. *Computational Statistics*, v. 27, p. 51–67, 2012.
- SOUZA, R. C.; MARCATO, A. L. M.; DIAS, B. H.; OLIVEIRA, F. L. C. Optimal operation of hydrothermal systems with hydrological scenario generation through bootstrap and periodic autoregressive models. *European Journal of Operational Research*, v. 222, p. 606–615, 2012.
- STARTZ, R. Binomial Autoregressive Moving Average Models with an Application to U.S. Recessions. *Journal of Business & Economic Statistics*, v. 26:1, p. 1–8, 2012.
- TIAO, G. C.; GRUPE, M. R. Hidden periodic autoregressive-moving average models in time series data. *Biometrika*, v. 67, p. 365–373, 1980.
- TOMY, L.; JOSE, K. K. Generalized normal-Laplace AR Process. *Statistics and Probability Letters*, v. 79, p. 1615–1620, 2009.
- TROUTMAN, B. M. Some results in periodic autoregress. *Biometrika*, v. 66, p. 219–228, 1979.
- TSIOTAS, G. On the use of non-linear transformations in stochastic volatility models. *Statistical Methods and Applications*, v. 18, p. 555–583, 2009.



- WANG, C.; LI, W. K. On the Autopersistence Functions and the Autopersistence Graphs of Binary Autoregressive Time Series. *Journal of Time Series Analysis*, v. 32, p. 639–646, 2011.
- ZEGER, S. L. A regression model for time series of counts. *Biometrika*, v. 75, p. 621–629, 1988.
- ZHENG, T.; SONG, T. A realized stochastic volatility model with Box-Cox transformation. *Journal of Business & Economic Statistics*, v. 32, 2014.

# Apêndice A

## Algumas Considerações sobre Processo Estocástico

**Definição A.1** Um processo estocástico  $\{Z_t, t \in \mathbb{N}\}$  diz-se fortemente estacionário se e somente se,

$$F_{X_{t_1}, \dots, X_{t_n}}(x_1, \dots, x_n) = F_{X_{t_1+\tau}, \dots, X_{t_n+\tau}}(x_1, \dots, x_n),$$

para quaisquer  $t_k \in \mathbb{N}$ ,  $n \in \mathbb{N}$  e  $\tau \in \mathbb{R}$ .

**Definição A.2** Um processo estocástico  $\{Z_t, t \in \mathbb{N}\}$  diz-se fracamente estacionário ou estacionário se e somente se,

- $E(Z_t) = \mu_t = \mu$ , constante para todo tempo  $t \in \mathbb{N}$ ;
- $E(Z_t^2) < \infty$ , constante para todo tempo  $t \in \mathbb{N}$ ;
- $Cov(Z_{t_1}, Z_{t_2})$  é uma função de  $|t_1 - t_2|$ .

**Teorema A.1 (Representação de Wold)** Se  $\{Z_t, t \in \mathbb{N}\}$  é um processo estocástico estacionário então este pode ser representado por

$$\begin{aligned} z_t &= \psi_0 + \sum_{j=1}^{\infty} \psi_j a_{t-j} + a_t, \\ &= \Psi(B)a_t, \end{aligned}$$

em que  $\{a_t, t \in \mathbb{N}\}$  é um processo estocástico com  $E(a_t) = 0$ ,  $Var(a_t) = \sigma_a^2$  e  $E(a_t a_{t+\tau}) = 0$  para todo  $\tau \neq 0$ .

**Corolário A.1** Um processo estocástico  $\{Z_t, t \in \mathbb{N}\}$ , dado por  $Z_t = \Psi(B)a_t$  é fracamente estacionário se  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ , com  $\Psi(B) = \sum_{j=0}^{\infty} \psi_j(B)$  e  $\psi_0 = 1$ .

**Definição A.3** A representação do processo estocástico  $\{Z_t, t \in \mathbb{N}\}$  em função de valores passados  $Z_{t-j}$ ,  $j = 1, 2, \dots$ , é dada por

$$Z_t = - \sum_{j=1}^{\infty} \pi_j Z_{t-j} + a_t, \quad (\text{A.1})$$

ou seja

$$Z_t + \sum_{j=1}^{\infty} \pi_j Z_{t-j} = a_t, \quad (\text{A.2})$$

então definindo  $\Pi(B) = 1 + \pi_1 B + \pi_2 B + \dots = 1 + \sum_{j=1}^{\infty} \pi_j B^j$ , podemos representar  $Z_t$  por

$$\Pi(B)Z_t = a_t, \quad (\text{A.3})$$

para que as duas representações  $Z_t = \Psi(B)a_t$  e  $\Pi(B)Z_t = a_t$  sejam a única para o mesmo processo  $\{Z_t, t \in \mathbb{N}\}$  devemos assegurar que

$$Z_t = \Psi(B)\Pi(B)Z_t, \quad (\text{A.4})$$

ou seja,  $\Pi(B) = \Psi(B)^{-1}$ .

**Definição A.4** Um processo estocástico  $\{Z_t, t \in \mathbb{N}\}$ , dado por  $Z_t = \Psi(B)a_t$  é invertível se existe um polinômio  $\Pi(B) = \Psi^{-1}(B)$  tal que

$$\Pi(B)Z_t = a_t$$

e

$$\sum_{j=0}^{\infty} |\pi_j| < \infty \quad (\text{A.5})$$

sendo  $\pi_0 = 1$ .

# Apêndice B

## Algumas Demonstrações

Prova da Proposição 2.1.

Defina  $X$  e  $W$  duas variáveis aleatórias tais que  $X = \sigma X_1 X_2^{1/s} + \mu$  e  $W = X_2$ , disto temos que  $X_2 = W$  e  $X_1 = \frac{X - \mu}{\sigma W^{1/s}}$ . As primeiras derivadas de  $X_1$  e  $X_2$  em relação a  $X$  e  $W$  são

$$\frac{\partial X_1}{\partial X} = \frac{1}{\sigma W^{1/s}}, \quad \frac{\partial X_1}{\partial W} = -\frac{X - \mu}{s\sigma W^{1/s+1}}, \quad \frac{\partial X_2}{\partial X} = 0, \quad \frac{\partial X_2}{\partial W} = 1$$

com isso temos que o determinante do jacobiano da transformação,  $J$ , é  $\frac{1}{\sigma W^{1/s}}$ .

Do fato que  $X_1$  e  $X_2$  são independentes segue que a densidade conjunta de  $X_1$  e  $X_2$  é dada por

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= f_{X_1}(x_1) f_{X_2}(x_2) \\ &= \frac{1}{2} I_{(-1,1)}(x_1) \frac{1}{\Gamma(1 + 1/s)} x_2^{1/s} e^{-x_2}. \end{aligned}$$

A densidade conjunta de  $X$  e  $W$  é dada por

$$\begin{aligned} f_{X,W}(x, w) &= |J| f_{X_1}(x_1(x, w)) f_{X_2}(x_2(x, w)) \\ &= \frac{1}{2\sigma\Gamma(1 + 1/s)} I_{(-1,1)}\left(\frac{x - \mu}{\sigma w^{1/s}}\right) e^{-w}. \end{aligned}$$

Observando que  $-1 < \frac{x - \mu}{\sigma w^{1/s}} < 1$  então  $w > \left|\frac{x - \mu}{\sigma}\right|^s$ , com isso temos

$$f_{X,W}(x, w) = \frac{1}{2\sigma\Gamma(1 + 1/s)} I_{(|\frac{x - \mu}{\sigma}|^s, \infty)}(w) e^{-w}. \quad (\text{B.1})$$

Como o objetivo é a densidade de  $X$  então integramos a densidade conjunta, dada em (B.1), em  $W$ , ou seja,

$$\begin{aligned} f_X(x) &= \int f_{X,W}(x,w)dw \\ &= \int_{|\frac{x-\mu}{\sigma}|^s}^{\infty} \frac{1}{2\sigma\Gamma(1+1/s)} e^{-w} dw \\ &= \frac{1}{2\sigma\Gamma(1+1/s)} e^{-|\frac{x-\mu}{\sigma}|^s}. \end{aligned}$$

Usando a propriedade  $\Gamma(1+1/s) = \frac{1}{s}\Gamma(1/s)$  temos

$$f_X(x) = \frac{s}{2\sigma\Gamma(1/s)} e^{-|\frac{x-\mu}{\sigma}|^s}.$$

### Prova da Proposição 2.2.

$$\begin{aligned} \int_0^{\infty} f_{X|X_2}(x|x_2)f_{X_2}(x_2)dx_2 &= \int_0^{\infty} \frac{1}{2x_2^{1/s}\sigma} I_{(\mu-x_2^{1/s}\sigma, \mu+x_2^{1/s}\sigma)}(x) \frac{1}{\Gamma(1+1/s)} x_2^{1/s} \exp(-x_2) dx_2 \\ &= \frac{1}{2\sigma\Gamma(1+1/s)} \int_0^{\infty} I_{(\mu-x_2^{1/s}\sigma, \mu+x_2^{1/s}\sigma)}(x) \exp(-x_2) dx_2, \quad (\text{B.2}) \end{aligned}$$

observe que  $\mu - x_2^{1/s}\sigma < x < \mu + x_2^{1/s}\sigma$ , implicando em  $x_2 > \left|\frac{x-\mu}{s}\right|^s$ , logo podemos reescrever a equação (B.2) como

$$\int_0^{\infty} f_{X|X_2}(x|x_2)f_{X_2}(x_2)dx_2 = \frac{1}{2\sigma\Gamma(1+1/s)} \int_0^{\infty} I_{(|\frac{x-\mu}{\sigma}|^s, \infty)}(x_2) \exp(-x_2) dx_2, \quad (\text{B.3})$$

utilizando a propriedade  $\Gamma(1+1/s) = \frac{1}{s}\Gamma(1/s)$  e resolvendo a integral na equação (B.3) temos

$$\int_0^{\infty} f_{X|X_2}(x|x_2)f_{X_2}(x_2)dx_2 = \frac{s}{2\sigma\Gamma(1+1/s)} \exp\left(-\left|\frac{x-\mu}{\sigma}\right|^s\right) = f_X(x)$$

### Prova do Teorema 2.1.

Defina  $z_t = \mu_t + \nu_t$ , logo  $\nu_t = z_t - \mu_t$  e  $\nu_t$  são erros martingale (ou seja, a média

é zero e são não correlacionados). Reescrevendo a equação (2.8) temos

$$\begin{aligned}
\mu_t - \mathbf{x}'_t \boldsymbol{\beta} &= \sum_{j=1}^p \phi_j (z_{t-j} - \mathbf{x}'_{t-j} \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) \Rightarrow \\
z_t - \nu_t - \mathbf{x}'_t \boldsymbol{\beta} &= \sum_{j=1}^p \phi_j (z_{t-j} - \mathbf{x}'_{t-j} \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) \Rightarrow \\
z_t - \mathbf{x}'_t \boldsymbol{\beta} &= \sum_{j=1}^p \phi_j (z_{t-j} - \mathbf{x}'_{t-j} \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) + \nu_t \Rightarrow \\
w_t &= \sum_{j=1}^p \phi_j w_{t-j} + \sum_{j=1}^q \theta_j \nu_{t-j} + \nu_t, \tag{B.4}
\end{aligned}$$

sendo que  $w_t = z_t - \mathbf{x}'_t \boldsymbol{\beta}$ ,  $w_{t-1} = z_{t-1} - \mathbf{x}'_{t-1} \boldsymbol{\beta}$ ,  $\dots$ ,  $w_1 = z_1 - \mathbf{x}'_1 \boldsymbol{\beta}$ . Com isso, temos

$$\begin{aligned}
w_t &= \Phi(B)^{-1} \Theta(B) \nu_t, \\
&= \Psi(B) \nu_t,
\end{aligned}$$

sendo  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ,  $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  e  $\Psi(B) = \Phi(B)^{-1} \Theta(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$  (desde que  $\sum_{i=0}^{+\infty} |\psi_i| < +\infty$ ). Portanto

$$E\{Z_t\} = E\{\mathbf{x}'_t \boldsymbol{\beta} + W_t\} = \mathbf{x}'_t \boldsymbol{\beta} + E\{W_t\} = \mathbf{x}'_t \boldsymbol{\beta}.$$

Adotando  $\mathbf{x}'_t \boldsymbol{\beta} = \mathbf{x}'_{t-1} \boldsymbol{\beta} = \dots = \mathbf{x}'_1 \boldsymbol{\beta} = \beta$ ,  $z_t = \mu_t + \nu_t$  e  $w_t = z_t - \beta$ ,  $w_{t-1} = z_{t-1} - \beta$ ,  $\dots$ ,  $w_1 = z_1 - \beta$ , conseguimos reescrever a equação (2.8) como a equação (B.4). Logo,  $E\{Z_t\} = E\{\beta + W_t\} = \beta$ . Observe que ainda podemos reescrever a equação (2.8) na forma

$$\begin{aligned}
z_t &= \beta - \sum_{j=1}^p \phi_j \beta + \sum_{j=1}^p \phi_j z_{t-j} + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) + \nu_t \Rightarrow \\
z_t &= (1 - \sum_{j=1}^p \phi_j) \beta + \sum_{j=1}^p \phi_j z_{t-j} + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) + \nu_t \Rightarrow \\
z_t &= \beta_0 + \sum_{j=1}^p \phi_j z_{t-j} + \sum_{j=1}^q \theta_j (z_{t-j} - \mu_{t-j}) + \nu_t,
\end{aligned}$$

ou seja,  $\beta_0 = (1 - \sum_{j=1}^p \phi_j) \beta$ , assim

$$E\{Z_t\} = E\{\beta + W_t\} = \beta = \frac{\beta_0}{1 - \sum_{j=1}^p \phi_j}.$$

Prova do Teorema 2.2.

Sejam  $w_t = z_t - \mathbf{x}'_t \boldsymbol{\beta}$ ,  $\Phi(B)$ ,  $\Theta(B)$  e  $\Psi(B)$  como definidos no teorema anterior e no enunciado, observamos que

$$\begin{aligned}
 Var(Z_t) &= Var(\mathbf{x}'_t \boldsymbol{\beta} + W_t) = Var(W_t) = E(W_t^2) \\
 &= E(\Psi(B)\nu_t\Psi(B)\nu_t) \\
 &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j E(\nu_{t-i}\nu_{t-j}) \\
 &= \sum_{i=0}^{\infty} \psi_i^2 E(\nu_{t-i}^2), \tag{B.5}
 \end{aligned}$$

mas

$$\begin{aligned}
 E(\nu_t^2) &= E(E(\nu_t^2 | \mathfrak{F}_{t-1})) = E(E((Z_t - \mu_t)^2 | \mathfrak{F}_{t-1})) \\
 &= E(Var(Z_t | \mathfrak{F}_{t-1})) \\
 &= E\left(\frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)}\right) = \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)},
 \end{aligned}$$

logo reescrevendo a equação (B.5) temos

$$Var(Z_t) = \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)} \sum_{i=0}^{\infty} \psi_i^2,$$

e no caso particular em que  $p = q = 1$ , observemos que

$$\Psi(B) = \Phi(B)^{-1} \Theta(B) \Rightarrow \Phi(B) \Psi(B) = \Theta(B),$$

logo

$$(1 - \phi_1 B)(\psi_0 + \psi_1 B + \psi_2 B + \dots) = (1 + \theta_1 B),$$

portanto  $\psi_0 = 1$ ,  $\psi_1 = \phi_1 + \theta_1$ ,  $\psi_2 = \phi_1(\phi_1 + \theta_1)$ ,  $\dots$ ,  $\psi_n = \phi_1^{n-1}(\phi_1 + \theta_1)$ , disto temos

$$\begin{aligned}
 Var(Z_t) &= \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)} \left(1 + \sum_{i=1}^{\infty} \phi_1^{i-1} (\phi_1 + \theta_1)\right) \\
 &= \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)} \left(1 + \frac{\phi_1 + \theta_1}{1 - \phi_1}\right).
 \end{aligned}$$