

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

**ABORDAGEM CLÁSSICA E BAYESIANA PARA OS MODELOS
DE SÉRIES TEMPORAIS DA FAMÍLIA GARMA COM
APLICAÇÕES PARA DADOS CONTÍNUOS**

Marcos Henrique Cascone

São Carlos/SP

Maio/2011

Abordagem clássica e bayesiana para os modelos de séries
temporais da família GARMA com aplicações para dados
contínuos

Marcos Henrique Cascone

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

**Abordagem clássica e bayesiana para os modelos de
séries temporais da família GARMA com aplicações
para dados contínuos**

Marcos Henrique Cascone

Orientador: Prof. Dr. Marinho Gomes de Andrade Filho

Dissertação apresentada ao Programa de
Pós-Graduação de Estatística, para obten-
ção do título de Mestre em Estatística.

São Carlos/SP

Maio/2011

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

C336ac

Cascone, Marcos Henrique.

Abordagem clássica e bayesiana para os modelos de séries temporais da família GARMA com aplicações para dados contínuos / Marcos Henrique Cascone. -- São Carlos : UFSCar, 2011.

82 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2011.

1. Análise de séries temporais. 2. Modelos estatísticos. 3. Distribuição normal. 4. Distribuição inversa Gaussiana. 5. Distribuição gama. 6. Modelos GARMA. I. Título.

CDD: 519.55 (20^a)

Marcos Henrique Cascone

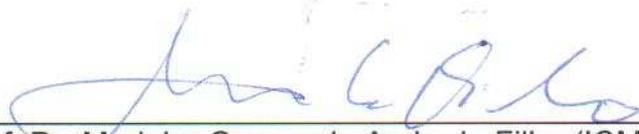
**ABORDAGEM CLÁSSICA E BAYESIANA PARA OS MODELOS DE
SÉRIES TEMPORAIS DA FAMÍLIA GARMA COM APLICAÇÕES
PARA DADOS CONTÍNUOS**

Dissertação apresentada à Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de Mestre em Estatística.

Aprovada em 24 de março de 2011.

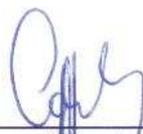
BANCA EXAMINADORA

Presidente



Prof. Dr. Marinho Gomes de Andrade Filho (ICMC-USP/Orientador)

1º Examinador



Prof. Dr. Carlos Alberto Ribeiro Diniz (DEs-UFSCar)

2º Examinador



Prof. Dr. Valderio Anselmo Reisen (UFES)

"Não basta ensinar ao homem uma especialidade, porque se tornará, assim, uma máquina utilizável e não uma personalidade. É necessário que adquira um sentimento, um senso prático daquilo que vale a pena ser empreendido, daquilo que é belo, do que é moralmente correto."

Albert Einstein

Agradecimentos

Primeiramente agradeço aos meus pais Edison e Iris que me deram não somente a vida, mas principalmente a minha educação e condições de estudo e apoio incondicional, os mais profundos agradecimentos. A minha irmã, Fernanda, que sempre esteve carinhosamente presente, pelas críticas e força para superar obstáculos todos os dias na difícil tarefa de estudar. Sem vocês, nada disso teria sido possível. A Deus, por mais essa oportunidade concedida e concluída.

A minha namorada, amiga e companheira Adriana, por todo o incentivo concedido, pela paciência, por aguentar momentos de ansiedade e estresse e por estar sempre ao meu lado, estudando, auxiliando e apoiando para que eu continuasse e concluísse mais uma etapa de nossas vidas que vamos construindo juntos.

Ao Professor Doutor Marinho Gomes de Andrade Filho, orientador da dissertação, pela paciente tolerância e oportunidade que a mim foi confiada e por estar sempre tão solícito para dirimir toda e qualquer dúvida, pela orientação e incentivo na elaboração e condução deste trabalho.

A todos amigos e amigas, agradeço sinceramente pela amizade, pela preocupação e por todos os momentos divertidos.

Aos professores Luis Aparecido Milan e Ricardo Sandes Ehlers, membros da banca do exame de qualificação, pelas sugestões feitas.

Aos funcionários do Departamento de Estatística, em especial a Isabel, secretária do departamento, por propiciar sempre boas condições de trabalho.

A todas as pessoas que, direta ou indiretamente, contribuíram para a realização dessa dissertação.

Finalmente, agradeço à CAPES pelo auxílio concedido para este trabalho.

Resumo

Neste trabalho, o objetivo foi analisar no contexto clássico e bayesiano, o modelo GARMA com três distribuições contínuas: Gaussiana (Normal), Inversa Gaussiana e Gama, e também o desempenho e a qualidade do ajuste dos modelos de interesse, bem como o desempenho dos percentis de cobertura para eles. Para o estudo clássico foi considerado os estimadores de máxima verossimilhança e por meio de simulação verificou-se a consistência, o viés e o erro quadrático médio dos mesmos. Para a abordagem bayesiana é proposta uma distribuição a priori não informativa para os parâmetros dos modelos resultando em uma distribuição a posteriori, o qual a partir daí pode-se encontrar as estimativas bayesianas para os parâmetros, sendo que este estudo ainda não foi encontrado na literatura. Com isso pode-se observar que a inferência bayesiana mostrou boa eficiência no processo de análise da série, o que pode ser comprovado também com a última etapa do trabalho. Esta, consiste na análise de um conjunto de dados reais correspondente a taxa de casos de tuberculose na região metropolitana de São Paulo. Os resultados mostram que, tanto o estudo clássico quanto o bayesiano, são capazes de descrever bem o comportamento da série.

Palavras-chave: Modelo GARMA, Distribuição Normal, Distribuição Inversa Gaussiana, Distribuição Gama, Inferência Clássica, Inferência Bayesiana.

Abstract

In this work, the aim was to analyze in the classic and bayesian context, the GARMA model with three different continuous distributions: Gaussian, Inverse Gaussian and Gamma. We analyzed the performance and the goodness of fit of the three models, as well as the performance of the coverage percentile. In the classic analyze we consider the maximum likelihood estimator and by simulation study, we verified the consistency, the bias and de mean square error of the models. To the bayesian approach we proposed a non-informative prior distribution for the parameters of the model, resulting in a posterior distribution, which we found the bayesian estimatives for the parameters. This study still was not found in the literature. So, we can observe that the bayesian inference showed a good quality in the analysis of the serie, which can be comprove with the last section of this work. This, consist in the analyze of a real data set corresponding in the rate of tuberculosis cases in metropolitan area of Sao Paulo. The results show that, either the classical and bayesian approach, are good alternatives to describe the behavior of the real time serie.

Keywords: *GARMA Model, Gaussian Distribution, Inverse Gaussian Distribution, Gamma Distribution, Classic Inference, Bayesian Inference.*

Sumário

1	Introdução	10
1.1	Processos lineares estacionários	11
1.2	Modelos ARMA(p, q)	13
1.3	Modelos ARIMA(p, d, q)	14
1.4	Organização do texto	15
2	Modelos GARMA(p, q)	17
2.1	Introdução	17
2.2	Modelos Lineares Generalizados	18
2.3	Especificação do Modelo GARMA	19
2.3.1	Casos particulares do Modelo GARMA	21
3	Ajuste do modelo: Abordagem Clássica para Modelos GARMA	24
3.1	Função de verossimilhança	24
3.2	Método Escore de Fisher	25
3.3	Seleção e avaliação de modelos	28
3.4	Análise de resíduos	29
3.5	Previsão para os Modelos GARMA	29

4	Inferência Bayesiana	31
4.1	Especificação de distribuições a priori	33
4.1.1	Priori não Informativa	33
4.1.2	Priori informativas	35
4.2	Abordagem Bayesiana para o modelo GARMA	36
4.3	Métodos MCMC	41
4.4	Seleção de modelos	44
4.5	Previsão	45
5	Aplicações	47
5.1	Aplicações com dados simulados	47
5.1.1	Estudo de simulação	48
5.1.2	Ajustes clássicos e bayesianos para séries simuladas	55
5.2	Aplicação com dados reais	60
5.2.1	Taxa de casos de tuberculose na região metropolitana de SP.	60
5.2.2	Método Escore de Fisher	61
5.2.3	Abordagem bayesiana	67
5.2.4	Comentários finais sobre as aplicações	71
6	Conclusões e considerações finais	72
	Referências	74
A	Geração dos dados	77
B	Modelos GARMA para séries financeiras	79
B.0.5	Modelo GARCH	80

B.0.6	O modelo GARMA GARCH	81
-------	--------------------------------	----

Capítulo 1

Introdução

A análise de séries temporais é uma área da estatística dedicada ao estudo de observações que apresentam dependência no tempo. Uma série temporal pode ser definida como um conjunto de observações de uma variável dispostas sequencialmente no tempo. Esses tipos de dados surgem nas mais variadas áreas de aplicação como, por exemplo, em finanças, em econometria, em meteorologia, nas ciências sociais, na física, na geofísica entre outras.

Na literatura encontram-se vários modelos de séries temporais, dos quais destacamos o modelo autorregressivo e de médias móveis de ordem (p, q) , pois o mesmo é muito utilizado quando, encontramos na prática, séries estacionárias as quais queremos representar por um modelo linear e com um número não muito grande de parâmetros.

Se considerarmos uma extensão dos modelos gaussianos autorregressivos e de médias móveis (ARMA) para uma série não gaussiana, teremos que garantir que a distribuição da série temporal utilizada pertença a uma família de distribuição tratável, como por exemplo a família exponencial. Com isso poderemos escrever o modelo de forma muito semelhante aos Modelos Lineares Generalizados (MLGs), com algumas diferenças que serão citadas no decorrer do texto. Essa classe de modelos inclui distribuições da família exponencial como por exemplo: Poisson, Binomial e Binomial Negativa, do lado discreto, Normal, Gama e Inversa Gaussiana do lado contínuo.

Os MLGs foram introduzidos inicialmente por Nelder & Wedderburn (1972), sendo

baseados em distribuições que são modelos de dispersão exponencial. Os MLGs podem ser estendidos para séries temporais e vários autores apresentam casos especiais de MLGs para análise de série temporal. Veja, por exemplo, Li (1994); Benjamin *et al.* (2003); Fokianos & Kedem (2004), nos quais os modelos GARMA Poisson, binomial logística, binomial negativa e gama são discutidos e aplicados a conjunto de dados de série temporal real.

A utilização de transformações amplia mais as possibilidades de utilização dos modelos GARMA por permitirem que séries cujas distribuições são diferente das distribuições já citadas sejam modeladas após uma transformação adequada. Em um trabalho recente, Cordeiro e Andrade (2007) desenvolvem uma nova classe de modelos lineares generalizados transformados que estendem os modelos de Box e Cox (1964) para os modelos da família exponencial generalizados ARMA, resultando nos modelos TGARMA (Transformed Generalized ARMA).

Os modelos TGARMA, além de permitir a inclusão de co-variáveis, permitem que sejam considerados vários modelos da família exponencial para modelar a função distribuição de probabilidade condicional da variável resposta transformada. Essa flexibilidade pode trazer ganhos significativos na modelagem e precisão das previsões de séries financeiras.

A utilização dos modelos GARMA (e/ou TGARMA) para modelagem de séries financeiras é uma abordagem nova no sentido de que se mostra que os modelos GARCH são casos particulares dos modelos GARMA (desde que com erros normais), os quais por sua vez, permitem abordar a modelagem de séries financeiras considerando modelos da família exponencial. A utilização de transformações (Modelos TGARMA) amplia mais as possibilidades de utilização dos modelos GARMA, pois permitem que séries cujas distribuições apresentam caudas pesadas e assimétricas sejam modeladas, após uma transformação adequada. Sendo assim a utilização de modelos TGARMA para séries financeiras mostra-se como um tema novo e bastante promissor.

1.1 Processos lineares estacionários

Definição 1.1 *Um processo estocástico $\{y_t\}$, $\forall t$ é estacionário no sentido forte (ou estritamente estacionário) se $F_{y_{t_1}, \dots, y_{t_n}}(y_1, \dots, y_n) = F_{y_{t_1+\tau}, \dots, y_{t_n+\tau}}(y_1, \dots, y_n)$, para todo $t_1 <$*

$t_2 < \dots \leq t_n \in T$, $n \geq 1$ e para todo $\tau \in T$, ou seja, se todas as distribuições finito-dimensionais $F_{y_{t_1}, \dots, y_{t_n}}(y_1, \dots, y_n) = P(y_{t_1}(w) \leq y_1, \dots, y_{t_n}(w) \leq y_n)$, (função distribuição de probabilidade conjunta), permanecem as mesmas sob translações do tempo.

A estacionariedade em seu sentido forte é, em geral, de difícil verificação. Uma forma menos restrita para se definir estacionariedade é a que segue.

Definição 1.2 Um processo estocástico $\{y_t\}$, $\forall t$, é fracamente estacionário (ou estacionário de segunda ordem) se $E(y_t) = \mu = \text{constante}$, $\text{var}(y_t) = \sigma^2 = \text{constante}$ e $\text{cov}(y_t, y_{t+k}) = \gamma(k)$ (covariância entre y_t e y_{t+k} é uma função que só depende de k), para todo t e $k \in \mathbb{Z}$.

Observação 1.1 A estacionariedade em seu sentido fraco não implica na estacionariedade forte (Brockwell & Davis, 2002).

A Definição 1.3 tem por objetivo introduzir o conceito de ruído branco para uso em processos lineares.

Definição 1.3 Se $\{a_t, t = 1, 2, \dots\}$ são variáveis aleatórias independentes e identicamente distribuídas com média zero e variância constante σ_a^2 , então $\{a_t\}$ recebe o nome de ruído branco.

Definição 1.4 Uma série temporal $\{y_t\}$ é um processo linear se tem representação

$$y_t = \sum_{j=-\infty}^{\infty} \psi_j a_{t-j}, \quad (1.1)$$

para todo t , em que a_t é ruído branco com média 0 e variância constante σ^2 e ψ_j é uma sequência de constantes com $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

A equação (1.1) pode ser escrita em termos do operador de retrocesso B , ou seja,

$$y_t = \psi(B)a_t, \quad (1.2)$$

com $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B_j$ (Brockwell & Davis, 2002).

Para definir a estacionariedade para processos lineares precisaremos do Teorema de Wald.

Teorema 1.1 (Teorema de Wald) *Se $\{y_t, t = 1, 2, \dots\}$ é um processo linear, então existe um processo estocástico $\{a_t, t = 1, 2, \dots\}$ tal que*

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

com $\psi_0 = 1$, $E(a_t) = 0$, $\forall t$ e $E(a_t a_s) = 0$ para todo $s \neq t$ e $E(a_t^2) = \sigma_a^2$, em que σ_a^2 denota a variância de a_t .

Além disso, pode-se dizer que $E(y_t) = \mu$ se a série $\sum_{j=1}^n \psi_j < \infty$.

Diz-se que o processo linear dado em (1.1) é estacionário no sentido fraco se as raízes do polinômio $\psi(B) = 0$ estão dentro ou na borda do círculo de raio unitário. Se a_t são processos independentes e identicamente distribuídos $N(0, \sigma_a^2)$, esta condição das raízes de $\psi(B) = 0$ garantem a estacionariedade no sentido forte para o processo.

1.2 Modelos ARMA(p, q)

Uma metodologia bastante utilizada na análise de modelos paramétricos é conhecida como abordagem de Box e Jenkins, que consiste em ajustar modelos autorregressivos e de médias móveis ARMA(p, q) a um conjunto de dados. Será dado aqui, somente uma introdução do modelo e para uma leitura com maior quantidade de detalhes, ver Box *et al.* (2008).

Considere $\{y_t, t = 1, 2, \dots\}$ uma série temporal. $\{y_t\}$ é um modelo ARMA(p, q) com média c se $\{y_t\}$ é estacionário e para todo t ,

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = c + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q},$$

em que $a_t \sim N(0, \sigma^2)$ e os polinômios $\phi(B)$ e $\theta(B)$ são, respectivamente, o polinômio autorregressivo estacionário de ordem p e o polinômio de médias móveis de ordem q .

Observação 1.2 B é um operador (chamado operador de retrocesso) tal que $B^k y_t = y_{t-k}$.

Estacionariedade para o modelo ARMA(p, q)

Seja $\{y_t\}$ um processo ARMA(p, q). $\{y_t\}$ é estacionário se existem constantes ψ_j com $\sum_{j=0}^{\infty} |\psi_j| < \infty$ e

$$y_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

para todo t . Essa condição é equivalente a $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \neq 0$ para todo $|B| < 1$.

Invertibilidade para o modelo ARMA(p, q)

Seja $\{y_t\}$ um processo ARMA(p, q). $\{y_t\}$ é invertível se existem constantes π_j com $\sum_{j=0}^{\infty} |\pi_j| < \infty$ e

$$a_t = \sum_{j=0}^{\infty} \pi_j y_{t-j}$$

para todo t . Essa condição é equivalente a $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots - \pi_p B^p \neq 0$ para todo $|B| < 1$.

Do exposto acima, conclui-se que o processo ARMA(p, q) será estacionário se as raízes do polinômio autorregressivo $\phi(B) = 0$ estiverem fora do círculo unitário, e será invertível se as raízes do polinômio de médias móveis $\theta(B) = 0$ estiverem também fora do círculo de raio unitário.

1.3 Modelos ARIMA(p, d, q)

O Modelo ARMA(p, q) é em geral utilizado para descrever séries estacionárias, sem qualquer tipo de tendência e que se desenvolvem no tempo ao redor de uma média constante. Porém, na prática, grande parte das séries encontradas são não-estacionárias, como por exemplo, séries econômicas, séries de consumo, séries de preços, etc. que, quando diferenciadas, tornam-se estacionárias.

Para ilustrar o escrito anterior, suponha y_t uma série não-estacionária. Calculando uma diferença na série

$$y_t - y_{t-1} = (1 - B)y_t = \Delta y_t = w_t,$$

tem-se que a série w_t torna-se estacionária.

Ainda, se $w_t = \Delta^d y_t$ for estacionária, pode-se representar w_t por um modelo ARMA(p, q), ou seja,

$$\phi(B)w_t = \theta(B)a_t.$$

Se w_t for uma diferença de y_t , então y_t segue um modelo autorregressivo, *integrado*, de médias móveis, ou ARIMA, dado por

$$\phi(B)\Delta^d y_t = \theta(B)a_t \quad (1.3)$$

de ordem (p, d, q) , escrevendo-se da forma ARIMA(p, d, q). A ordem d é o número de diferenças feitas na série para torná-la estacionária e as ordens p e q funcionam como no modelo ARMA, respectivamente, as ordens dos polinômios autorregressivos e de médias móveis.

Uma desvantagem do uso dos modelos ARIMA na forma (1.3) é que a cada diferenciação da série, perde-se uma informação, ou seja, se forem feitas 3 diferenças para tornar a série estacionária, o tamanho reduzirá de n para $n - 3$, por exemplo. Dado que essas informações podem ser relevantes para o estudo, essa perda pode ser significativa. Outro ponto importante é que nem sempre tomar diferenças na série pode ser suficiente para se obter a estacionariedade, como acontece em muitas séries financeiras. Para isso, pode ser necessário fazer alguma transformação antes da diferenciação para que se possa alcançar a estacionariedade desejada.

Para um estudo mais detalhado sobre os modelos ARIMA ver Box *et al.* (2008).

1.4 Organização do texto

No Capítulo 2 deste trabalho é apresentado um resumo teórico do modelo GARMMA, bem como alguns de seus casos particulares com distribuições da classe contínua. No Capítulo 3 é apresentada a abordagem clássica do modelo tratado, com a estimação dos parâmetros, construção de intervalos de confiança, seleção de modelos, análise de resíduos e previsão. No Capítulo 4 é proposta uma abordagem bayesiana para o modelo, com a eliciação de distribuições a priori, estimação de parâmetros, seleção de modelos e previsão.

No Capítulo 5 serão apresentadas algumas aplicações com o modelo GARMA, entre elas um estudo de simulação e uma aplicação com dados reais utilizando os métodos descritos nos Capítulos 3 e 4. Finalmente, no Capítulo 6 são sugeridos os caminhos a serem seguidos para dar continuidade ao trabalho aqui iniciado.

Capítulo 2

Modelos GARMA(p, q)

2.1 Introdução

Quando se trata de séries com dependência temporal é importante lembrar que nem sempre as séries apresentaram ruídos normais e, diversas vezes, não são estacionárias. Esses fatos acabam impossibilitando o uso dos modelos ARMA de Box e Jenkins. Um exemplo clássico de série que não é aconselhável ser ajustada por um modelo ARMA já foi mencionado no capítulo anterior, que são séries financeiras. Pode-se agregar a essas, séries cujas distribuições têm caudas mais pesadas ou até mesmo assimetria, as quais o modelo ARMA também não seria eficientemente ajustado.

Os modelos GARMA são uma extensão dos Modelos Lineares Generalizados de McCullagh & Nelder (1989) para situações de dados dependentes, caracterizando-se pela adição de um termo extra ao preditor linear, no qual este passa a incorporar termos autorregressivos (AR) e de médias móveis (MA). Dessa forma, pode-se utilizar distribuições contínuas assimétricas, como por exemplo, Gama e Inversa Gaussiana, para o caso de assimetria nos dados. Pode-se então expandir a quantidade de séries que podem ser tratadas com o modelo GARMA, desde que as mesmas tenham distribuições que pertençam a família exponencial.

Tem-se também os casos de séries temporais de dados de contagem, que são registros da frequência relativa de ocorrência de certos eventos em sucessivos intervalos de tempo,

geralmente igualmente espaçados. Podem ser citados como exemplos o número mensal de sinistros de carros associados a uma seguradora, o número diário de internações hospitalares em uma cidade em consequência de problemas cardio-respiratórios, entre outros. Os modelos apropriados para dados de contagem devem utilizar distribuições de probabilidade cujo domínio seja formado apenas pelos números inteiros não-negativos como é o caso das distribuições Poisson, binomial e binomial negativa, as quais podem ser utilizados nos modelos GARMA, já que essas distribuições pertencem também à família exponencial. Porém o foco deste trabalho está somente nas distribuições contínuas, por isso não será abordado nada a respeito de séries de dados de contagem.

Com o objetivo de estudar o modelo GARMA, será feita uma pequena introdução dos MLG e sobre família exponencial biparamétrica.

2.2 Modelos Lineares Generalizados

Muitas das distribuições conhecidas podem ser reunidas em uma família denominada exponencial de distribuições, como por exemplo, a distribuição normal, poisson, binomial, gama, entre várias outras (Cordeiro *et al.*, 1995). A importância da família exponencial de distribuições teve maior destaque, na área dos modelos de regressão, a partir do trabalho pioneiro de Nelder & Wedderburn (1972) que definiram os modelos lineares generalizados.

Dizemos que Y tem distribuição na família exponencial biparamétrica se

$$f_Y(y|\theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\},$$

sendo θ e $\phi > 0$ respectivamente os parâmetros canônico e de dispersão e $b(\cdot)$ e $c(\cdot)$ funções conhecidas. A média e variância são dadas por

$$E(Y) = \mu = b'(\theta) \quad \text{e} \quad Var(Y) = \phi b''(\theta) = \phi V(\mu),$$

em que $V(\mu) = V_t = b''(\theta)$ é denominada função de variância e é expressa em termos da média μ . Essa desempenha um papel importante na família exponencial, uma vez que a mesma caracteriza a distribuição.

Os MLGs podem ser usados quando se tem uma observação de $\mathbf{Y} = (y_1, \dots, y_n)$ com elementos independentes e um vetor de médias $\boldsymbol{\mu}$, com $E(Y_i) = \mu_i = \sum_{j=1}^p x_{ij}\beta_j$, $i = 1, 2, \dots, n$, sendo que $p < n$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ é o vetor de coeficientes desconhecidos e $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ são os vetores de covariáveis.

Em forma matricial, $\boldsymbol{\mu}_{n \times 1} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1}$, em que \mathbf{X} é uma matriz de posto igual a p , chamada de matriz modelo. $\mathbf{X}\boldsymbol{\beta}$ é chamado preditor linear e é denotado por $\boldsymbol{\eta}$, isto é, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$.

Um MLG envolve três componentes:

- **Componente aleatório:** é representado por um conjunto de variáveis aleatórias independentes pertencentes a uma mesma distribuição, que faça parte da família exponencial, com médias $E(Y_i) = \mu_i$, $i = 1, \dots, n$.
- **Componente sistemático:** as variáveis explicativas $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ fornecem um preditor linear $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$.
- **Função de Ligação:** é uma função que relaciona o componente aleatório ao componente sistemático, ou seja, a média ao preditor linear, ou seja,

$$g(\mu_i) = \eta_i$$

em que $g(\cdot)$ uma função monótona e diferenciável.

Com os conceitos iniciais de MLG definidos, daremos início à explanação dos modelos GARMA.

2.3 Especificação do Modelo GARMA

Considere \mathbf{H}_t o conjunto de informações passadas das observações, médias e covariáveis, ou seja,

$$\mathbf{H}_t = \{\mathbf{x}_t, \dots, \mathbf{x}_1, y_{t-1}, \dots, y_1, \mu_{t-1}, \dots, \mu_1\}.$$

Em um modelo GARMA, a distribuição condicional de cada observação y_t para cada $t = 1, \dots, n$, dado \mathbf{H}_t , é considerada pertencer à família exponencial se

$$f(y_t|\mathbf{H}_t) = \exp \left\{ \frac{y_t \theta_t - b(\theta_t)}{\varphi} + c(y_t, \varphi) \right\}, \quad (2.1)$$

em que θ_t e φ são os parâmetros canônico e de dispersão, respectivamente. As funções $b(\cdot)$ e $c(\cdot)$ definem uma família exponencial específica e \mathbf{x} é o vetor de variáveis explicativas.

Segundo Cordeiro & Andrade (2009), para distribuições contínuas, (2.1) é uma família exponencial completa biparamétrica com parâmetros canônicos $1/\varphi$ e θ/φ e possui a seguinte decomposição

$$c(x, \varphi) = \frac{1}{\varphi} a(x) + d(\varphi) + d_1(x). \quad (2.2)$$

A equação (2.2) engloba os mais importantes MGLs contínuos: os modelos normal, gama e inverso Gaussiano.

No entanto, a diferença principal entre os MLG e os modelos GARMA está no preditor, pois neste último é incluído um componente adicional, τ_t , referindo-se a parte relativa aos termos autorregressivos e de médias móveis, além de incluir no preditor valores passados das variáveis explicativas do modelo.

Assim, temos que

$$g(\mu_t) = \eta_t = \mathbf{x}_t \boldsymbol{\beta} + \tau_t, \quad (2.3)$$

em que

$$\tau_t = \sum_{j=1}^p \phi_j \mathcal{A}(y_{t-j}, x_{t-j}, \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j \mathcal{M}(y_{t-j}, \mu_{t-j}),$$

sendo que \mathcal{A} e \mathcal{M} são funções que representam os termos autorregressivos e de médias móveis, com $\boldsymbol{\phi}' = (\phi_1, \dots, \phi_p)$ e $\boldsymbol{\theta}' = (\theta_1, \dots, \theta_p)$ os parâmetros autorregressivos e de médias móveis respectivamente.

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

$$g(\mu_t) = \eta_t = \mathbf{x}_t\boldsymbol{\beta} + \underbrace{\sum_{j=1}^p \phi_j \{g(y_{t-j}) - \mathbf{x}_{t-j}\boldsymbol{\beta}\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}) - \eta_{t-j}\}}_{\tau_t}, \quad (2.4)$$

sendo este mais flexível e parcimonioso. As equações (2.1) e (2.4) definem um modelo GARMA.

Tem-se então a possibilidade de usar diferentes distribuições, desde que a mesma pertença à família exponencial. O uso dessas distribuições resulta em diferentes casos particulares do modelo GARMA, sendo que alguns serão citados a seguir.

2.3.1 Casos particulares do Modelo GARMA

Uma das vantagens do uso dos modelos GARMA para séries temporais está no fato de abrir o leque de opções de conjuntos de dados a trabalhar. Nesta seção serão estudados três casos particulares do modelo GARMA, atribuindo as distribuições Normal, Gama e Inversa Gaussiana, sendo elas as distribuições contínuas mais comuns, e também um caso particular do modelo GARMA Gama (Apêndice B).

Modelo GARMA Gaussiano

Seja $\{y_t\}$, $t = 1, \dots, n$ uma série temporal e considere uma distribuição normal para $y_t | \mathbf{H}_t$. Então

$$f(y_t | \mathbf{H}_t) = \exp \left\{ \frac{1}{\sigma^2} \left(y_t \mu_t - \frac{\mu_t^2}{2} \right) - \frac{1}{2} \log(2\pi\sigma^2) - \frac{y_t^2}{2\sigma^2} \right\}$$

com $y_t \in \mathbb{R}$, $\mu_t \in \mathbb{R}$ e $\sigma > 0$, em que

$$\theta_t = \mu_t, \quad \varphi = \sigma^2, \quad b(\theta_t) = \frac{\theta_t^2}{2} \quad \text{e} \quad c(y_t, \sigma^2) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{y_t^2}{2\sigma^2}.$$

A função de variância e a função de ligação canônica para esse modelo são dadas respectivamente por $V_t = 1$ e $g(\mu_t) = \mu_t$. Então o preditor linear é

$$\mu_t = \eta_t = \mathbf{x}_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{y_{t-j} - \mathbf{x}_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \theta_j \{y_{t-j} - \mu_{t-j}\}.$$

Neste trabalho será considerado somente os casos em que $\boldsymbol{\beta} = \beta_0$. Então

$$\mu_t = \beta_0 + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \theta_j \{y_{t-j} - \mu_{t-j}\}.$$

Como $y_t | \mathbf{H}_t \sim N(\mu_t, \sigma^2)$, então $y_t - \mu_t \sim N(0, \sigma^2)$. Denotando-se por $a_t = y_t - \mu_t$, tem-se que $\mu_t = y_t - a_t$. Então,

$$\begin{aligned} y_t - a_t &= \beta_0 + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \theta_j a_{t-j} \\ y_t &= \beta_0 + \sum_{j=1}^p \phi_j y_{t-j} + a_t + \sum_{j=1}^q \theta_j a_{t-j}. \end{aligned} \quad (2.5)$$

Observe que o modelo (2.5) é na verdade um modelo ARMA(p,q). Essa é uma peculiaridade do modelo GARMA Gaussiano, sendo este capaz de fornecer o modelo ARMA usual.

Modelo GARMA Inversa Gaussiana

Seja $\{y_t\}$, $t = 1, \dots, n$ uma série temporal. Suponha agora que $y_t | \mathbf{H}_t$ tenha distribuição Inversa Gaussiana. Então

$$f(y_t | \mathbf{H}_t) = \left(\frac{1}{2\pi\sigma^2 y_t^3} \right)^{(1/2)} \exp \left\{ -\frac{(y_t - \mu_t)^2}{2\mu_t^2 \sigma^2 y_t} \right\}, \quad (2.6)$$

com $y_t > 0$, $\mu_t > 0$ e $\sigma > 0$.

A função de densidade (2.6) escrita na forma da família exponencial é dada por

$$f(y_t | \mathbf{H}_t) = \exp \left\{ \frac{1}{\sigma^2} \left(-\frac{y_t}{2\mu_t^2} + \frac{1}{\mu_t} \right) - \frac{1}{2} \left[\frac{1}{y_t \sigma^2} + \log(2\pi\sigma^2 y_t^3) \right] \right\}, \quad (2.7)$$

em que

$$\theta_t = -\frac{1}{2\mu_t^2}, \quad \varphi = \sigma^2, \quad b(\theta_t) = -(-2\theta_t)^{1/2} \text{ e } c(y_t, \sigma^2) = \frac{1}{2} \left[\frac{1}{y_t \sigma^2} + \log(2\pi\sigma^2 y_t^3) \right].$$

A função de ligação canônica para este modelo é a função $g(\mu_t) = 1/\mu_t^2$ e a função de variância é $V_t = \mu_t^3$. Portanto o preditor linear é dado por

$$\frac{1}{\mu_t^2} = \eta_t = \mathbf{x}_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \left\{ \frac{1}{y_{t-j}^2} - \mathbf{x}_{t-j} \boldsymbol{\beta} \right\} + \sum_{j=1}^q \theta_j \left\{ \frac{1}{y_{t-j}^2} - \frac{1}{\mu_{t-j}^2} \right\}. \quad (2.8)$$

Logo, as equações (2.7) e (2.8) definem o modelo GARMA Inversa Gaussiana.

Modelo GARMA Gama

Considere $\{y_t\}$, $t = 1, \dots, n$ uma série temporal e suponha que $y_t | \mathbf{H}_t$ tenha distribuição Gama. Seja a reparametrização $\nu = \alpha$ e $\mu_t = \alpha\beta \Rightarrow \beta = \mu_t/\nu$. Então

$$f(y_t | \mathbf{H}_t) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu_t} \right)^\nu y_t^{(\nu-1)} \exp \left\{ -\frac{y_t \nu}{\mu_t} \right\},$$

com $y_t > 0$, $\mu_t > 0$ e $\nu_t > 0$.

Na forma da família exponencial, é escrita como

$$f(y_t | \mathbf{H}_t) = \exp \left\{ \frac{1}{\nu} \left(-\frac{y_t}{\mu_t} + \log \frac{1}{\mu_t} \right) + \frac{1}{\nu} \log \left(\frac{y_t}{\nu} \right) - \log \Gamma \left(\frac{1}{\nu} \right) - \log(y_t) \right\}. \quad (2.9)$$

A função de variância é $V_t = \mu_t^2$ e a função de ligação canônica para este modelo é a função recíproca $g(\mu_t) = 1/\mu_t$. Dessa maneira, o preditor linear é

$$\frac{1}{\mu_t} = \eta_t = \mathbf{x}_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \left\{ \frac{1}{y_{t-j}} - \mathbf{x}_{t-j} \boldsymbol{\beta} \right\} + \sum_{j=1}^q \theta_j \left\{ \frac{1}{y_{t-j}} - \frac{1}{\mu_{t-j}} \right\}, \quad (2.10)$$

em que

$$\theta_t = -\frac{1}{\mu_t}, \quad \varphi = \nu, \quad b(\theta_t) = -\log(-\theta_t) \quad \text{e} \quad c(y_t, \nu) = \frac{1}{\nu} \log \left(\frac{y_t}{\nu} \right) - \log \Gamma \left(\frac{1}{\nu} \right) - \log(y_t).$$

Logo, as equações (2.9) e (2.10) definem o modelo Gama-GARMA.

Agora especificados os modelos que vão ser explorados, serão feitas as inferências para todos esses modelos, como seguem nos próximos capítulos.

Capítulo 3

Ajuste do modelo: Abordagem Clássica para Modelos GARMA

3.1 Função de verossimilhança

Seja $\{y_t\}$, $t = 1, \dots, n$ uma série temporal e suponha que a variável resposta y_t satisfaça as condições dos modelos GARMA dadas em (2.1) e (2.4). O principal objetivo da análise desses modelos é fazer inferência sobre os seus parâmetros, dados por $\boldsymbol{\gamma}' = (\boldsymbol{\beta}', \boldsymbol{\phi}', \boldsymbol{\theta}')$, em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)'$, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)'$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$ mais o parâmetro de dispersão, chamado de φ . O procedimento utilizado será o de máxima verossimilhança aproximada, o qual é baseado na distribuição de probabilidade da variável resposta dadas as respostas passadas e as informações passadas das covariáveis.

A função de verossimilhança aproximada do vetor de parâmetros $\boldsymbol{\gamma}$ e do escalar φ , condicionada às r primeiras observações, é

$$\mathcal{L}(\boldsymbol{\gamma}, \varphi) = \prod_{t=r+1}^n f(y_t | \mathbf{H}_t),$$

com $r = \max(p, q)$.

A log-verossimilhança aproximada para os parâmetros do modelo é dada por

$$\ell(\boldsymbol{\gamma}, \varphi) = \frac{1}{\varphi} \sum_{t=r+1}^n \{y_t \theta_t\} + \sum_{t=r+1}^n c(y_t, \varphi).$$

3.2 Método Escore de Fisher

Considere as matrizes A e B definidas a seguir, cujos elementos são funções dos parâmetros do modelo.

$$\mathbf{A} = \begin{bmatrix} g(y_r) - x_r \boldsymbol{\beta} & \cdots & g(y_{r+1-p}) - x_{r+1-p} \boldsymbol{\beta} \\ g(y_{r+1}) - x_{r+1} \boldsymbol{\beta} & \cdots & g(y_{r+2-p}) - x_{r+2-p} \boldsymbol{\beta} \\ \vdots & \ddots & \vdots \\ g(y_{n-1}) - x_{n-1} \boldsymbol{\beta} & \cdots & g(y_{n-p}) - x_{n-p} \boldsymbol{\beta} \end{bmatrix}_{(n-r) \times p},$$

$$\mathbf{B} = \begin{bmatrix} g(y_r) - \eta_r & \cdots & g(y_{r+1-q}) - \eta_{r+1-q} \\ g(y_{r+1}) - \eta_{r+1} & \cdots & g(y_{r+2-q}) - \eta_{r+2-q} \\ \vdots & \ddots & \vdots \\ g(y_{n-1}) - \eta_{n-1} & \cdots & g(y_{n-q}) - \eta_{n-q} \end{bmatrix}_{(n-r) \times q}.$$

O componente sistemático do modelo GARMA será escrito em forma matricial, correspondente às observações y_{r+1}, \dots, y_n , dado por

$$\boldsymbol{\eta} = \begin{bmatrix} \mathbf{X} & \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\phi} \\ \boldsymbol{\theta} \end{bmatrix},$$

em que \mathbf{X} é a matriz local formada pelas linhas \mathbf{x}_t , $t = r + 1, \dots, n$. Pode-se também denotar o componente sistemático reduzindo-o a apenas $\boldsymbol{\eta} = \mathbf{M}\boldsymbol{\gamma}$, em que $\mathbf{M} = [\mathbf{X} \ \mathbf{A} \ \mathbf{B}]$ é a matriz modelo de ordem $(n - r) \times (m + p + q)$.

Seja $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\eta}} = \widehat{\mathbf{M}}\hat{\boldsymbol{\gamma}}$ e $\hat{\varphi}$ os estimadores de máxima verossimilhança (EMV) de $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ e φ respectivamente. Uma característica importante no processo de estimação dos parâmetros é que a estimação do vetor de parâmetros $\boldsymbol{\gamma}$ não depende de φ . Os EMVs $\hat{\boldsymbol{\gamma}}$ podem ser obtidos através do ajuste dos modelos (2.1) e (2.4) para y pelo método de mínimos quadrados ponderados reiterados, dado por

$$\hat{\boldsymbol{\gamma}}^{(k+1)} = \left(\widehat{\mathbf{M}}^{(k)} \widehat{\mathbf{W}}^{(k)} \widehat{\mathbf{M}}^{(k)} \right)^{-1} \widehat{\mathbf{M}}^{(k)} \widehat{\mathbf{W}}^{(k)} \hat{z}^{(k)}, \quad (3.1)$$

com k sendo o passo da iteração, $\widehat{\mathbf{W}} = \text{diag}\{w_{r+1}, \dots, w_n\}$ é a matriz de pesos, com $w_t = V_t^{-1} \left(\frac{dg(\mu_t)}{d\mu_t} \right)^{-2}$ e a variável z_t é dada por

$$z_t = \eta_t + (y_t - \mu_t) \frac{dg(\mu_t)}{d\mu_t}.$$

O procedimento iterativo usando a equação (3.1) parte de um valor inicial $\hat{\gamma}^{(1)}$ e calcula $\widehat{\mathbf{M}}^{(1)}$, $\widehat{\mathbf{W}}^{(1)}$ e $\hat{z}^{(1)}$, usados para obter a próxima estimativa $\hat{\gamma}^{(2)}$ em (3.1). Este procedimento é realizado até que a convergência seja observada.

As estimativas $\hat{\eta}_t$, para $t = r + 1, \dots, n$ são obtidas de

$$\hat{\eta}_t = \mathbf{x}_t \hat{\boldsymbol{\beta}} + \sum_{j=1}^p \hat{\phi}_j \left\{ g(y_{t-j}) - \mathbf{x}_{t-j} \hat{\boldsymbol{\beta}} \right\} + \sum_{j=1}^q \hat{\theta}_j \left\{ g(y_{t-j}) - \hat{\eta}_{t-j} \right\},$$

e as médias ajustadas são obtidas através da inversa da função g , isto é, $\hat{\mu}_t = g^{-1}(\hat{\eta}_t)$, com $t = r + 1, \dots, n$.

Para estimar φ , basta derivar a log-verossimilhança em relação a φ . Fazendo isso obtemos

$$\hat{\varphi}^2 \sum_{t=r+1}^n \frac{dc(y_t, \varphi)}{d\varphi} \Big|_{\varphi=\hat{\varphi}} = \sum_{t=r+1}^n \left\{ y_t \hat{\theta}_t - b(\hat{\theta}_t) \right\},$$

em que $\hat{\theta}_t = q(g^{-1}(\mathbf{x}_t \hat{\boldsymbol{\beta}}))$ e dada a função de variância $V(x)$, podemos obter $q(x) = \int V^{-1}(x)$.

A função desvio D para o modelo GARMA é

$$D = 2 \sum_{t=r+1}^n \left\{ y_t q(y_t) - b(q(y_t)) \right\} - 2 \sum_{t=r+1}^n \left\{ y_t q(\hat{\mu}_t) - b(q(\hat{\mu}_t)) \right\}$$

e a estimativa de máxima verossimilhança parcial de $\hat{\varphi}$ pode ser escrita como

$$\hat{\varphi}^2 \sum_{t=r+1}^n \frac{dc(y_t, \varphi)}{d\varphi} \Big|_{\varphi=\hat{\varphi}} = \sum_{t=r+1}^n e(y_t) - \frac{D}{2}, \quad (3.2)$$

com $e(x) = xq(x) - b(q(x))$. A equação (3.2) é em geral não linear exceto para os casos dos modelos Normal e Inversa Gaussiana.

No caso de uma família exponencial biparamétrica, a decomposição da função $c(y_t, \varphi)$ dada em (2.2) resulta na equação abaixo para $\hat{\varphi}$,

$$(n - r) \hat{\varphi}^2 d'(\hat{\varphi}) = \sum_{t=r+1}^n t(y_t) - \frac{D}{2}, \quad (3.3)$$

com $t(x) = xq(x) - b(q(x)) + a(x)$.

A Tabela 3.1 exhibe as funções $d(x)$ e $t(x)$ necessárias para calcular $\hat{\varphi}$ em (3.3).

Tabela 3.1: Funções para o cálculo de $\hat{\varphi}$.

Modelo	$d(x)$	$t(x)$
Normal	$-\frac{1}{2} \log(x)$	0
Gama	$-\frac{\log(x)}{x} - \log\left(\Gamma\left(\frac{1}{x}\right)\right)$	-1
Inversa Gaussiana	$-\frac{1}{2} \log(x)$	0

No entanto, para os modelos Normal e Inverso Gaussiano pode-se estimar φ de uma maneira mais direta, através da equação

$$\hat{\varphi} = \frac{D}{n - r}.$$

Para o modelo Gama, quando trata-se de valores pequenos de φ , pode-se utilizar uma solução aproximada que é dada por

$$\hat{\varphi} = \frac{2D}{(n - r) \left\{ 1 + \left(1 + \frac{2D}{3(n - r)} \right)^{1/2} \right\}}. \quad (3.4)$$

Caso contrário, a estimação de φ requer o uso de um processo iterativo, utilizando as funções da Tabela 3.1 na equação (3.3). Por se tratar de um processo iterativo, em grande parte dos casos usa-se como valor inicial a solução aproximada dada em (3.4) e um método bastante utilizado é o método iterativo de Newton-Raphson, dado pela expressão

$$\hat{\varphi}^{(k)} = \hat{\varphi}^{(k-1)} - \frac{f(\hat{\varphi}^{(k-1)})}{f'(\hat{\varphi}^{(k-1)})},$$

em que k é a iteração do processo e $f'(\hat{\varphi})$ é a derivada da função f de (3.3). A cada iteração a solução é avaliada até que se obtenha a convergência desejada. Ver Cordeiro & Andrade (2009) para obter maiores detalhes das construções feitas anteriormente.

Pode-se também fazer inferência com respeito aos parâmetros φ e γ , isto é, intervalos de confiança, testes de hipóteses como por exemplo o teste da razão de verossimilhança, análise de desvio, análise de resíduos e diagnósticos (Cordeiro & Andrade, 2009).

A matriz aproximada de variância e covariâncias de $\hat{\gamma}$ é dada por $\hat{\varphi}(\mathbf{M}'\mathbf{W}\mathbf{M})^{-1}$. A variância de $\hat{\varphi}$, para o modelo Normal e Inverso Gaussiano é $Var(\hat{\varphi}) = 2\hat{\varphi}^2/n$ e para o modelo Gama é dada por $Var(\hat{\varphi}) = n^{-1}\hat{\varphi}^4 \{\psi(\hat{\varphi}^{-1}) - \hat{\varphi}\}^{-1}$, em que $\psi(\cdot)$ é a função trigamma.

3.3 Seleção e avaliação de modelos

Para comparar modelos, pode-se utilizar os critérios AIC e BIC. O critério AIC (Akaike Information Criterion) (Akaike, 1974) é definido por

$$AIC = -2\ell(\boldsymbol{\gamma}) + 2k,$$

em que $\ell(\boldsymbol{\gamma})$ representa o máximo da função de log-verossimilhança e k é o número de parâmetros do modelo. Esse critério é baseado em considerações frequentistas de eficiência assintótica.

O segundo método em questão é o critério BIC (Bayesian Information Criterion) (Schwarz, 1978) definido por

$$BIC = -2\ell(\boldsymbol{\gamma}) + k \log(n),$$

com os parâmetros sendo os mesmos definidos para o critério AIC, porém introduzido $\log(n)$, em que n é o número de observações da série.

Os dois critérios apresentados, AIC e BIC, têm como objetivo introduzir a complexidade do modelo no critério de seleção, pois são critérios que “penalizam” a verossimilhança. Através do uso destas métricas o modelo mais indicado é aquele que apresentar menor valor para AIC e BIC.

Por fim, pode-se avaliar os modelos ajustados através do Erro Quadrático Médio (EQM) e do Erro Percentil Absoluto Médio (EPAM) que são dados respectivamente por

$$EQM = \frac{100\%}{n\hat{\sigma}_y^2} \sum_{t=1}^n (y_t - \hat{\mu}_t)^2 \quad (3.5)$$

e

$$EPAM = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{\mu}_t}{y_t} \right|, \quad (3.6)$$

em que $\hat{\sigma}_y^2$ é a variância amostral de Y .

3.4 Análise de resíduos

Após a seleção de modelos é necessário fazer um diagnóstico para a variável resposta e esse pode ser feito através da análise de resíduos. Os resíduos nada mais são do que a distância entre cada componente da variável resposta para sua média ajustada correspondente, ou seja, para cada $t = r + 1, \dots, n$,

$$r_t = y_t - \hat{\mu}_t,$$

em que $\hat{\mu}_t$ são as médias estimadas para cada variável resposta.

Nos modelos clássicos de regressão, após todo o procedimento de seleção de modelos e ajuste, a análise de resíduos é feita considerando os seguintes pressupostos: os resíduos são independentes (não correlacionados) com média 0 e variância constante σ^2 . Se algum desses pressupostos não forem válidos, significa que o modelo ajustado não é adequado ao conjunto de dados.

No âmbito dos MLG, a hipótese de normalidade não pode ser garantida, já que diferentes distribuições podem ser usadas desde que pertençam à família exponencial, diferentemente do modelo clássico de regressão onde a distribuição é restrita à Normal.

Diante desse fato, a análise de resíduos feita no presente trabalho irá apenas avaliar se os resíduos são não correlacionados com média 0 e variância constante. Para isso serão utilizados os gráficos de dispersão, de autocorrelação e autocorrelação parcial.

3.5 Previsão para os Modelos GARMA

Após a seleção de modelos e análise de resíduos, o último passo da análise a ser feito é a previsão. O interesse é prever um valor Y_{t+h} , $h > 1$, supondo que as observações $\dots, Y_{t-2}, Y_{t-1}, Y_t$ são conhecidas. Estas últimas são chamadas de *origem* das previsões.

A previsão de origem t e *horizonte* h será denotada de $\hat{Y}_t(h)$.

Nos Modelos GARMA a previsão é feita de maneira simplificada, através do preditor linear para cada modelo. Como todas as componentes do preditor são conhecidas, encontrar os valores da previsão se torna uma tarefa simples. Para $h = 1$, tem-se

$$\hat{\mu}_{t+1} = \hat{\alpha}_0 + \sum_{j=1}^p \hat{\phi}_j y_{t+1-j} + \sum_{j=1}^q \hat{\theta}_j \{y_{t+1-j} - \mu_{t+1-j}\}$$

para o modelo Gaussiano,

$$\hat{\mu}_{t+1} = \left(\hat{\alpha}_0 + \sum_{j=1}^p \hat{\phi}_j \frac{1}{y_{t+1-j}^2} + \sum_{j=1}^q \hat{\theta}_j \left\{ \frac{1}{y_{t+1-j}^2} - \frac{1}{\mu_{t+1-j}^2} \right\} \right)^{-1/2}$$

para o modelo Inversa Gaussiana e

$$\hat{\mu}_{t+1} = \left(\hat{\alpha}_0 + \sum_{j=1}^p \hat{\phi}_j \frac{1}{y_{t+1-j}} + \sum_{j=1}^q \hat{\theta}_j \left\{ \frac{1}{y_{t+1-j}} - \frac{1}{\mu_{t+1-j}} \right\} \right)^{-1}$$

para o modelo Gama.

Entretanto, pode-se generalizar as equações para uma previsão h -passos a frente, ou seja,

$$\hat{\mu}_{t+h} = \hat{\alpha}_0 + \sum_{j=1}^p \hat{\phi}_j y_{t+h-j} + \sum_{j=1}^q \hat{\theta}_j \{y_{t+h-j} - \mu_{t+h-j}\}, \quad (3.7)$$

$$\hat{\mu}_{t+h} = \left(\hat{\alpha}_0 + \sum_{j=1}^p \hat{\phi}_j \frac{1}{y_{t+h-j}^2} + \sum_{j=1}^q \hat{\theta}_j \left\{ \frac{1}{y_{t+h-j}^2} - \frac{1}{\mu_{t+h-j}^2} \right\} \right)^{-1/2} \quad (3.8)$$

e

$$\hat{\mu}_{t+h} = \left(\hat{\alpha}_0 + \sum_{j=1}^p \hat{\phi}_j \frac{1}{y_{t+h-j}} + \sum_{j=1}^q \hat{\theta}_j \left\{ \frac{1}{y_{t+h-j}} - \frac{1}{\mu_{t+h-j}} \right\} \right)^{-1} \quad (3.9)$$

para os modelos Gaussiano, Inverso Gaussiano e Gama respectivamente.

No entanto, faz-se necessário o uso de algumas técnicas para avaliar a qualidade dos valores previstos, ou seja, se estão ou não coerentes com o conjunto de dados. Um método largamente utilizado na literatura é a retirada de valores finais da série, calcular esses valores através das equações (3.7)-(3.9) e compará-los com os valores retirados da amostra. A quantidade de observações retiradas vai depender do tamanho do passo da previsão. Por exemplo, se o passo for de tamanho 5 então deve-se retirar 5 observações da amostra. Dessa forma que será feita a previsão para o conjunto de dados reais na Subseção 5.2.2 do Capítulo 5.

Capítulo 4

Inferência Bayesiana

Nos últimos tempos houve um aumento da pesquisa no campo bayesiano refletido pelo número de artigos e livros publicados. Uma razão para esse crescimento é a disponibilidade de algoritmos computacionais usados no cálculo de integrais que são necessárias em uma análise bayesiana.

A metodologia bayesiana é um tipo de inferência estatística que descreve as incertezas sobre quantidades de interesse de forma probabilística. Incertezas são modificadas periodicamente após observações de novos dados ou resultados. A operação que avalia a medida destas incertezas é conhecida como operação bayesiana e é baseada no teorema de Bayes.

Suponhamos que temos interesse em uma quantidade desconhecido γ . A informação sobre γ é resumida por um modelo probabilístico $\pi(\gamma)$, ou seja, expressa a incerteza sobre γ antes de observarmos os dados. Essa informação pode ser aumentada conforme uma quantidade aleatória Y relacionada com γ é observada $\pi(y|\gamma)$. A ideia de que após observar $Y = \mathbf{y}$, a quantidade de informação sobre γ aumenta, é bastante intuitiva e o teorema de Bayes é a regra para quantificar este aumento de informação.

Seja $\mathbf{y} = (y_1, \dots, y_n)$ uma amostra aleatória obtidos de uma população que se distribui de acordo com a densidade $f(\mathbf{y}|\gamma)$, $\gamma = (\gamma_1, \dots, \gamma_k)'$. O procedimento de inferência Bayesiana é baseado na forma do teorema de Bayes, isto é,

$$\pi(\gamma|\mathbf{y}) = \frac{\pi(\gamma, \mathbf{y})}{\pi(\mathbf{y})} = \frac{\pi(\mathbf{y}|\gamma)\pi(\gamma)}{\pi(\mathbf{y})} = \frac{\pi(\mathbf{y}|\gamma)\pi(\gamma)}{\int \pi(\gamma, \mathbf{y})d\gamma}.$$

Observe que $\pi(\mathbf{y})$ não depende de $\boldsymbol{\gamma}$. Assim esse termo funciona como uma constante normalizadora de $\pi(\boldsymbol{\gamma}|\mathbf{y})$.

Para um valor fixo de \mathbf{y} a função $\pi(\mathbf{y}|\boldsymbol{\gamma}) = \mathcal{L}(\boldsymbol{\gamma}; \mathbf{y})$ fornece a verossimilhança de cada um dos possíveis valores de $\boldsymbol{\gamma}$, enquanto $\pi(\boldsymbol{\gamma})$ é chamada de distribuição a priori de $\boldsymbol{\gamma}$. Combinando-se a função de verossimilhança $\mathcal{L}(\boldsymbol{\gamma}; \mathbf{y})$ com a priori $\pi(\boldsymbol{\gamma})$ leva à distribuição a posteriori de $\boldsymbol{\gamma}$. A forma usual do teorema de Bayes é escrita como

$$\pi(\boldsymbol{\gamma}|\mathbf{y}) \propto \mathcal{L}(\boldsymbol{\gamma}; \mathbf{y})\pi(\boldsymbol{\gamma}).$$

De posse da distribuição a posteriori podemos examinar quaisquer aspectos de $\boldsymbol{\gamma}$, como a média, mediana, moda, variância, percentis, etc.

Caso tenhamos o interesse em alguma componente de $\boldsymbol{\gamma}$, digamos γ_i ($i = 1, \dots, k$), a distribuição marginal correspondente pode ser obtida integrando-se a distribuição $\pi(\boldsymbol{\gamma}|\mathbf{y})$, isto é,

$$\pi(\gamma_i|\mathbf{y}) = \int \mathcal{L}(\boldsymbol{\gamma}_{-i}; \mathbf{y})\pi(\boldsymbol{\gamma}_{-i})d\boldsymbol{\gamma}_{-i}$$

onde $\boldsymbol{\gamma}_{-i} = (\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_k)$, ou seja, refere-se na integração de todos os componentes exceto γ_i .

Resumindo, na inferência bayesiana os parâmetros $\boldsymbol{\gamma}$ são tratados como quantidades aleatórias e o modelo estatístico não é mais somente $\pi(\mathbf{y}|\boldsymbol{\gamma})$ e sim $\pi(\mathbf{y}, \boldsymbol{\gamma})$, que representa a distribuição conjunta dos dados \mathbf{y} e dos parâmetros $\boldsymbol{\gamma}$. As estimativas para $\boldsymbol{\gamma}$ não serão somente valores, mas sim uma distribuição de probabilidades. A distribuição a priori $\pi(\boldsymbol{\gamma})$ expressa a incerteza sobre $\boldsymbol{\gamma}$ antes de observarmos os dados \mathbf{y} que dependem dele. A distribuição $\pi(\boldsymbol{\gamma}|\mathbf{y})$ expressa a incerteza sobre $\boldsymbol{\gamma}$ depois de observarmos os dados \mathbf{y} que dependem dele é denominada de distribuição a posteriori. Finalmente, de posse de $\pi(\boldsymbol{\gamma}|\mathbf{y})$, podemos examinar qualquer aspecto de $\boldsymbol{\gamma}$, como por exemplo a média, a variância, os percentis, a probabilidade de assumir determinados valores, entre outros. Os conceitos de priori e posteriori são relativos aquela observação que está sendo considerada no momento.

Na estimação bayesiana usualmente emprega-se a média da distribuição a posteriori como estimativas de parâmetros. Isso tem justificativa teórica no contexto de Teoria de Decisão. Quando se considera a função de perda quadrática, a média da distribuição a

posteriori é o valor que minimiza o risco de Bayes. Por sua vez, considerando a perda absoluta, a mediana da distribuição é a que minimiza o risco de Bayes e, por fim, considerando a função perda $0 - 1$, a moda da distribuição é o valor que minimiza o risco de Bayes. Muitas vezes, como resultado final das análises, costuma-se apresentar um resumo descritivo das distribuições marginais a posteriori dos parâmetros de interesse, incluindo também os valores das médias, medianas e modas dessas distribuições, bem como seus gráficos. Isso é realmente bastante ilustrativo, ficando para o usuário a tarefa de escolher a estimativa ou função perda do parâmetro a ser considerada.

4.1 Especificação de distribuições a priori

A utilização de informação a priori em inferência Bayesiana requer a especificação de uma distribuição a priori para os parâmetros de interesse. Esta distribuição deve representar o conhecimento que se tem sobre esses parâmetros antes da realização do experimento. Nesta seção, serão discutidas diferentes formas de especificação da distribuição a priori (Paulino *et al.*, 2003).

4.1.1 Priori não Informativa

Quando se espera que os dados transmitam toda a informação necessária, ou seja, a informação dos dados seja dominante, são utilizadas distribuições a priori “vagas”, caracterizadas pela ausência total, ou quantidade mínima de informação. Aqui serão citadas duas dessas prioris, que são as duas mais utilizadas para refletir essa ausência de informação.

A primeira se trata da distribuição a priori uniforme, na qual a ideia é de que todos os possíveis valores de γ são igualmente prováveis, e daí vem a distribuição uniforme. A ideia então é fazer $\pi(\gamma) \propto k$, com γ variando no subconjunto a qual ele pertence. No entanto, se γ for um conjunto infinito, então a distribuição a priori é imprópria, ou seja,

$$\int \pi(\gamma) d\gamma = \infty.$$

Neste caso, é necessário comprovar se a distribuição a posteriori obtidas com densidades a priori impróprias são próprias antes de se fazer inferência. Uma outra observação que deve

ser feita também, é a propriedade da invariância a transformações, a qual a priori uniforme não satisfaz.

Outro tipo de priori não informativa é a proposta por Jeffreys, em 1961, a qual possui a propriedade da invariância a transformações um a um, mas geralmente se trata de uma priori imprópria.

Antes de definir a priori de Jeffreys propriamente dita, é necessário introduzir alguns conceitos como o de Informação esperada e observada de Fisher.

Definição 4.1 *Considere uma única observação Y com função densidade de probabilidade $\pi(\mathbf{y}|\boldsymbol{\gamma})$. A medida de informação esperada de Fisher de $\boldsymbol{\gamma}$ através de Y é definida como*

$$I(\boldsymbol{\gamma}) = E \left[\frac{\partial \log \pi(\mathbf{y}|\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \cdot \frac{\partial \log \pi(\mathbf{y}|\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \right].$$

Se $\boldsymbol{\gamma}$ for um vetor paramétrico, define-se então a matriz de informação esperada de Fisher de $\boldsymbol{\gamma}$ através de Y como

$$\mathbf{I}(\boldsymbol{\gamma}) = -E \left[\frac{\partial^2 \log \pi(\mathbf{y}|\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right].$$

Pode-se então dizer que $I(\boldsymbol{\gamma})$ é uma medida de informação global, enquanto que uma medida de informação local pode ser obtida quando não se calcula o valor esperado na definição acima. Então a medida de informação observada de Fisher, denotada por $J(\boldsymbol{\gamma})$, é definida como

$$J(\boldsymbol{\gamma}) = \left(\frac{\partial \log \pi(\mathbf{y}|\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \right) \left(\frac{\partial \log \pi(\mathbf{y}|\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \right)'$$

Com as definições feitas acima, pode-se definir a priori de Jeffreys.

Definição 4.2 *Seja uma observação Y com função densidade de probabilidade $\pi(\mathbf{y}|\boldsymbol{\gamma})$. A priori de Jeffreys tem função densidade de probabilidade dada por*

$$\pi(\boldsymbol{\gamma}) \propto [I(\boldsymbol{\gamma})]^{1/2}. \quad (4.1)$$

Se $\boldsymbol{\gamma}$ for um vetor paramétrico, então $\pi(\boldsymbol{\gamma}) \propto [\det \mathbf{I}(\boldsymbol{\gamma})]^{1/2}$.

Para mais detalhes sobre essas duas distribuições a priori não informativas, ver Paulino *et al.* (2003).

4.1.2 Priori informativas

Diferentemente do que acontece com as distribuições a priori definidas acima, quando se tem um conhecimento sobre a quantidade γ de interesse, pode-se definir uma família paramétrica de densidades e então a distribuição a priori será representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento. Esses parâmetros da família de distribuições a priori são chamados de hiperparâmetros, para distingui-los do parâmetro de interesse γ . Pode-se também definir distribuições para os hiperparâmetros, ou seja, definir distribuições a prioris para eles. Neste caso, essas distribuições serão chamadas de distribuições a priori Hierárquicas, mas não serão abordadas neste trabalho.

O uso das prioris conjugadas em geral facilita a análise, uma vez que a ideia é de que as distribuições a priori e a posteriori pertençam a mesma classe de distribuições.

Definição 4.3 *Se $F = \{p(\mathbf{y}|\gamma), \gamma \in \Gamma\}$ é uma classe de distribuições amostrais, então uma classe de distribuições P é conjugada a F se*

$$\forall p(\mathbf{y}|\gamma) \in F \text{ e } p(\gamma) \in P \Rightarrow p(\gamma|\mathbf{y}) \in P.$$

A desvantagem do uso dessas prioris está no fato de que nem sempre a distribuição utilizada representa bem a incerteza inicial, e a vantagem do uso está no fato de ser mais simples a análise decorrente.

A forma de obtenção de uma família de distribuições conjugadas pode ser feita da seguinte forma:

- i) Identificar a classe P de distribuições para γ tal que $\ell(\gamma; \mathbf{y})$ seja proporcional a um membro desta classe;
- ii) Verifique se P é fechada por amostragem, isto é, se $\forall p_1, p_2 \in P \exists k$ tal que $kp_1p_2 \in P$.

Se além disso existir uma constante k tal que $k^{-1} = \int \ell(\gamma; y) d\gamma < \infty$ e todo $p \in P$ é definido como $p(\gamma) = k\ell(\gamma; y)$ então P é a *família conjugada natural* ao modelo amostral gerador de $\ell(\gamma; y)$.

Existe também uma outra forma de conjugação de prioris, que são as prioris conjugadas na família exponencial, bastante utilizadas por incluir grande parte das distribuições contínuas e discretas. Para maiores detalhes desta forma de conjugação, ver Ehlers (2007).

4.2 Abordagem Bayesiana para o modelo GARMA

A função de verossimilhança para o modelo GARMA é dada por

$$\mathcal{L}(\mu_t|y_t, \mathbf{H}_t) = \exp \left\{ \frac{1}{\varphi} \sum_{t=r+1}^n y_t [q(\mu_t) - b(q(\mu_t))] + \sum_{t=r+1}^n c(y_t, \varphi) \right\}, \quad (4.2)$$

em que $r = \max(p, q)$.

A Tabela 4.1 dá as possíveis funções a considerar em (4.2) para os três diferentes modelos utilizados.

Tabela 4.1: Funções especiais.

Modelo	$q(\mu_t)$	$b(q(u_t))$	$c(y_t, \varphi)$	$g(\mu_t)$
Normal	μ_t	$\frac{\mu_t^2}{2}$	$-\frac{1}{2} \left\{ \frac{y_t^2}{\varphi} + \log(2\pi\varphi) \right\}$	μ_t
Gama	$-\frac{1}{\mu_t}$	$\log(\mu_t)$	$\frac{1}{\varphi} \log\left(\frac{y_t}{\varphi}\right) - \log(y_t) - \log \Gamma\left(\frac{1}{\varphi}\right)$	$\frac{1}{\mu_t}$
Inversa Gaussiana	$-\frac{1}{2\mu_t^2}$	$-\frac{1}{\mu_t}$	$-\frac{1}{2} \left\{ \log(2\pi\varphi y_t^3) + \frac{1}{\varphi y_t} \right\}$	$\frac{1}{\mu_t^2}$

A função $g(\mu_t)$ na Tabela 4.1 corresponde a função de ligação canônica.

Sem perda de generalidade, será considerado $\mathbf{x}_t\boldsymbol{\beta} = \alpha_0$. Então, pode-se reescrever (2.4) da forma

$$g(\mu_t) = \alpha_0 + \sum_{i=1}^p \phi_i g(y_{t-i}) + \sum_{j=1}^q \theta_j \{g(y_{t-j}) - g(\mu_{t-j})\}. \quad (4.3)$$

Priori conjugadas para o modelo GARMA(p,q)

Dado que a distribuição de y_t pertence a família exponencial, se torna bastante fácil o uso de distribuições a priori conjugadas para esta classe de modelos. Serão consideradas

distribuições Normais independentes para o vetor de parâmetros $\boldsymbol{\gamma}$, com média 0 e variância τ , e uma distribuição Gama a priori para o parâmetro de dispersão φ , ou seja,

$$\pi(\alpha_0|\tau) \propto \frac{1}{\sqrt{\tau}} \exp\left\{-\frac{\alpha_0^2}{2\tau}\right\} \quad (4.4)$$

$$\pi(\phi_i|\tau) \propto \frac{1}{\sqrt{\tau}} \exp\left\{-\frac{\phi_i^2}{2\tau}\right\}, \quad i = 1, \dots, p \quad (4.5)$$

$$\pi(\theta_j|\tau) \propto \frac{1}{\sqrt{\tau}} \exp\left\{-\frac{\theta_j^2}{2\tau}\right\}, \quad j = 1, \dots, q \quad (4.6)$$

e

$$\pi(\varphi|\alpha, \beta) \propto \varphi^{(\alpha-1)} \exp\left\{-\frac{\varphi}{\beta}\right\}. \quad (4.7)$$

Considerando, então, as equações (4.2), (4.4), (4.5), (4.6) e (4.7), a distribuição a posteriori para distribuições a priori conjugadas a família exponencial será dada por

$$\pi(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \varphi|y_t, \mathbf{H}_t) \propto \mathcal{L}(\mu_t|y_t, \mathbf{H}_t) \pi(\alpha_0) \prod_{i=1}^p \pi(\phi_i) \prod_{j=1}^q \pi(\theta_j) \pi(\varphi)$$

e então

$$\begin{aligned} \pi(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \varphi|y_t, \mathbf{H}_t) &\propto \exp\left\{\frac{1}{\varphi} \sum_{t=r+1}^n y_t [q(\mu_t) - b(q(\mu_t))] + \sum_{t=r+1}^n c(y_t, \varphi)\right\} \\ &\times \frac{1}{\sqrt{\tau}} \exp\left\{-\frac{\alpha_0^2}{2\tau}\right\} \prod_{i=1}^p \frac{1}{\sqrt{\tau}} \exp\left\{-\frac{\phi_i^2}{2\tau}\right\} \\ &\times \prod_{j=1}^q \frac{1}{\sqrt{\tau}} \exp\left\{-\frac{\theta_j^2}{2\tau}\right\} \varphi^{(\alpha-1)} \exp\left\{-\frac{\varphi}{\beta}\right\}. \end{aligned}$$

Priori de Jeffreys para o modelo GARMA(p,q)

Considere $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta})$ o vetor de parâmetros e φ o parâmetro de dispersão. A log-verossimilhança para o modelo é dada por

$$\ell(\boldsymbol{\gamma}, \varphi) = \frac{1}{\varphi} \sum_{t=r+1}^n y_t [q(\mu_t) - b(q(\mu_t))] + \sum_{t=r+1}^n c(y_t, \varphi).$$

A função Escore para as componentes de $\boldsymbol{\gamma}$ são obtidas através de

$$U_i = \frac{\partial \ell(\boldsymbol{\gamma}, \varphi)}{\partial \gamma_i}, \quad i = 1, \dots, (m + p + q),$$

em que m , p e q são os comprimentos dos vetores $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ e $\boldsymbol{\theta}$ respectivamente.

Os elementos da matriz de informação para $\boldsymbol{\gamma}$ são dados por

$$\kappa_{i,j} = -E \left[\frac{\partial^2 \ell(\boldsymbol{\gamma}, \varphi)}{\partial \gamma_i \partial \gamma_j} \right] = E \left[\frac{\partial \ell(\boldsymbol{\gamma}, \varphi)}{\partial \gamma_i} \cdot \frac{\partial \ell(\boldsymbol{\gamma}, \varphi)}{\partial \gamma_j} \right],$$

com $i, j = 1, \dots, (m + p + q)$, e portanto

$$\kappa_{i,j} = \frac{1}{\varphi} \sum_{t=r+1}^n w_t y_{ti} y_{tj},$$

em que $w_t = \frac{1}{V_t} \left(\frac{\partial \mu_t}{\partial \eta_t} \right)^2$.

Considerando que os parâmetros $\boldsymbol{\gamma}$ e φ são independentes, a matriz de informação conjunta $\mathbf{K}(\boldsymbol{\gamma}, \varphi)$ é diagonal em blocos. Assim,

$$\mathbf{K}(\boldsymbol{\gamma}, \varphi) = \begin{bmatrix} K_1(\boldsymbol{\gamma}, \varphi) & 0 \\ 0 & \kappa_2(\boldsymbol{\gamma}, \varphi) \end{bmatrix},$$

em que $\kappa_2(\boldsymbol{\gamma}, \varphi) = -E \left[\frac{\partial^2 \ell(\boldsymbol{\gamma}, \varphi)}{\partial \varphi^2} \right]$ é a informação com relação a φ . A matriz de informação $K_1(\boldsymbol{\gamma}, \varphi)$ se reduz a

$$K_1(\boldsymbol{\gamma}, \varphi) = \varphi^{-1} \mathbf{M}' \mathbf{W} \mathbf{M},$$

com $\mathbf{W} = \text{diag}\{w_{r+1}, \dots, w_n\}$.

Calculando a segunda derivada de $\ell(\boldsymbol{\gamma}, \varphi)$ com relação a φ e calculando a esperança, tem-se que

$$E \left[\frac{\partial^2 \ell(\boldsymbol{\gamma}, \varphi)}{\partial \varphi^2} \right] = \frac{2}{\varphi^3} \sum_{t=r+1}^n y_t [q(\mu_t) - b(q(\mu_t))] + E \left[\frac{\partial^2 c(y_t, \varphi)}{\partial \varphi^2} \right]. \quad (4.8)$$

Utilizando (2.2), obtém-se

$$\frac{\partial c(y_t, \varphi)}{\partial \varphi} = -\frac{1}{\varphi^2} a(y_t) + d'(\varphi) \quad \text{e} \quad \frac{\partial^2 c(y_t, \varphi)}{\partial \varphi^2} = \frac{2}{\varphi^3} a(y_t) + d''(\varphi)$$

e então a equação (4.8) é escrita como

$$E \left[\frac{\partial^2 \ell(\boldsymbol{\gamma}, \varphi)}{\partial \varphi^2} \right] = \frac{2}{\varphi^3} \sum_{t=r+1}^n y_t [q(\mu_t) - b(q(\mu_t))] + \frac{2}{\varphi^3} \sum_{t=r+1}^n E [a(y_t)] + nd''(\varphi). \quad (4.9)$$

Sob as condições de regularidade, $a(y_t)$ pode ser calculado por

$$E \left[\frac{\partial \ell(\boldsymbol{\gamma}, \varphi)}{\partial \varphi} \right] = -\frac{1}{\varphi^2} \sum_{t=r+1}^n y_t [q(\mu_t) - b(q(\mu_t))] - \frac{1}{\varphi^2} \sum_{t=r+1}^n E [a(y_t)] + nd'(\varphi) = 0,$$

que resulta em

$$\sum_{t=r+1}^n E[a(y_t)] = n\varphi^2 d'(\varphi) - \sum_{t=r+1}^n y_t [q(\mu_t) - b(q(\mu_t))]. \quad (4.10)$$

Logo, usando (4.10) em (4.9), a matriz de informação para φ será dada por

$$\kappa_2(\boldsymbol{\gamma}, \varphi) = -n\varphi^{-1} \{2d'(\varphi) + \varphi d''(\varphi)\}.$$

Assim a matriz de informação conjunta $\mathbf{K}(\boldsymbol{\gamma}, \varphi)$ será dada por

$$\mathbf{K}(\boldsymbol{\gamma}, \varphi) = \begin{bmatrix} \varphi^{-1} \mathbf{M}' \mathbf{W} \mathbf{M} & 0 \\ 0 & n\varphi^{-1} v(\varphi) \end{bmatrix},$$

com $v(\varphi) = -\{2d'(\varphi) + \varphi d''(\varphi)\}$ e as funções $d(\varphi)$ e $v(\varphi)$ são dadas na Tabela 4.2.

Tabela 4.2: Funções $d(\varphi)$ e $v(\varphi)$.

Modelo	$d(\varphi)$	$v(\varphi)$
Normal	$-\frac{1}{2} \log(\varphi)$	$\frac{1}{2\varphi}$
Gama	$-\frac{\log(\varphi)}{\varphi} - \log \Gamma\left(\frac{1}{\varphi}\right)$	$-\frac{1}{\varphi^2} + \frac{1}{\varphi^3} \psi'\left(\frac{1}{\varphi}\right)$
Inversa Gaussiana	$-\frac{1}{2} \log(\varphi)$	$\frac{1}{2\varphi}$

A priori de Jeffreys é dada por (4.1), então $\pi(\boldsymbol{\gamma}, \varphi) \propto [\det(\mathbf{K}(\boldsymbol{\gamma}, \varphi))]^{1/2}$. Assim, a distribuição a priori conjunta de Jeffreys é dada por

$$\pi(\boldsymbol{\gamma}, \varphi) \propto \varphi^{-(p+1)/2} [v(\varphi)]^{1/2} \{\det(\mathbf{M}' \mathbf{W} \mathbf{M})\}^{1/2}.$$

A distribuição a posteriori pode ser obtida, então, por

$$\begin{aligned} \pi(\boldsymbol{\gamma}, \varphi | y_t, \mathbf{H}_t) &\propto \varphi^{-(p+1)/2} [v(\varphi)]^{1/2} \{\det(\mathbf{M}' \mathbf{W} \mathbf{M})\}^{1/2} \\ &\exp \left\{ \frac{1}{\varphi} \sum_{t=r+1}^n y_t [q(\mu_t) - b(q(\mu_t))] + \sum_{t=r+1}^n c(y_t, \varphi) \right\}. \end{aligned}$$

Priori não informativa para o modelo GARMA(p,q)

Um tipo de priori não informativa são as constantes, ou seja, $\pi(\boldsymbol{\gamma}) \propto \mathbf{c}$, em que \mathbf{c} é um vetor de constantes com tamanho $(m + p + q)$. Para a abordagem bayesiana deste problema, são consideradas distribuições a priori constantes com $c = 1$.

Para o parâmetro de dispersão presente em cada uma das distribuições, será utilizada priori não informativa, ou seja, $\pi(\varphi) \propto 1/\varphi$, $\varphi > 0$.

Pode-se então escrever a distribuição a posteriori como

$$\pi(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \varphi | y_t, \mathbf{H}_t) \propto \mathcal{L}(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \varphi | y_t, \mathbf{H}_t) \pi(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}) \pi(\varphi).$$

Já que $\pi(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta})$ é constante, então a distribuição a posteriori será dada por

$$\pi(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \varphi | y_t, \mathbf{H}_t) \propto \frac{1}{\varphi} \exp \left\{ \frac{1}{\varphi} \sum_{t=r+1}^n y_t [q(\mu_t) - b(q(\mu_t))] + \sum_{t=r+1}^n c(y_t, \varphi) \right\}. \quad (4.11)$$

Nas simulações e no ajuste a dados reais deste trabalho foram utilizadas distribuições a posteriori não informativa para os parâmetros, como dado em (4.11). Para uma melhor visualização dessas distribuições, serão explicitadas suas formas para cada submodelo trabalhado.

Modelo GARMA Gaussiano

A distribuição a posteriori para o modelo GARMA Gaussiano é dada por

$$\begin{aligned} \pi(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2 | y_t, \mathbf{H}_t) \propto & \frac{1}{\sigma^2} \exp \left\{ \frac{1}{\sigma^2} \sum_{t=r+1}^n \left[y_t \left(\alpha_0 + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \theta_j (y_{t-j} - \mu_{t-j}) \right) \right. \right. \\ & - \frac{1}{2} \left(\alpha_0 + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \theta_j (y_{t-j} - \mu_{t-j}) \right)^2 \\ & \left. \left. - \sum_{t=r+1}^n \frac{1}{2} \left(\frac{y_t^2}{\sigma^2} + 2\pi\sigma^2 \right) \right] \right\}. \end{aligned}$$

Modelo GARMA Inverso Gaussiano

A distribuição a posteriori para o modelo GARMA Inverso Gaussiano é definida da seguinte forma:

$$\begin{aligned} \pi(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2 | y_t, \mathbf{H}_t) &\propto \frac{1}{\sigma^2} \exp \left\{ \frac{1}{\sigma^2} \sum_{t=r+1}^n \left[-\frac{y_t}{2} \left(\alpha_0 + \sum_{j=1}^p \phi_j \frac{1}{y_{t-j}^2} + \sum_{j=1}^q \theta_j \left(\frac{1}{y_{t-j}^2} - \frac{1}{\mu_{t-j}^2} \right) \right) \right. \right. \\ &\quad \left. \left. + \left(\alpha_0 + \sum_{j=1}^p \phi_j \frac{1}{y_{t-j}^2} + \sum_{j=1}^q \theta_j \left(\frac{1}{y_{t-j}^2} - \frac{1}{\mu_{t-j}^2} \right) \right) \right]^{-2} \right. \\ &\quad \left. - \frac{1}{2} \sum_{t=r+1}^n \left[\frac{1}{y_t \sigma^2} + \log(2\pi \sigma^2 y_t^3) \right] \right\}. \end{aligned}$$

Modelo GARMA Gama

Por fim, a distribuição a posteriori para o modelo GARMA Gama é dada por

$$\begin{aligned} \pi(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \nu | y_t, \mathbf{H}_t) &\propto \frac{1}{\nu} \exp \left\{ \frac{1}{\nu} \sum_{t=r+1}^n \left[-y_t \left(\alpha_0 + \sum_{j=1}^p \phi_j \frac{1}{y_{t-j}} + \sum_{j=1}^q \theta_j \left(\frac{1}{y_{t-j}} - \frac{1}{\mu_{t-j}} \right) \right) \right. \right. \\ &\quad \left. \left. - \log \left(\alpha_0 + \sum_{j=1}^p \phi_j \frac{1}{y_{t-j}} + \sum_{j=1}^q \theta_j \left(\frac{1}{y_{t-j}} - \frac{1}{\mu_{t-j}} \right) \right) \right] \right. \\ &\quad \left. + \frac{1}{\nu} \sum_{t=r+1}^n \log \left(\frac{y_t}{\nu} \right) - \log \Gamma \left(\frac{1}{\nu} \right) - \sum_{t=r+1}^n \log y_t \right\}. \end{aligned}$$

Utilizando as distribuições a posteriori explicitadas, foi utilizado método de simulação MCMC, em particular o algoritmo de Metropolis, para obter as densidades marginais para α_0 , $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ e φ . Sumários das densidades a posteriori são apresentados com estimativas de Monte Carlo calculadas com as amostras geradas.

4.3 Métodos MCMC

Os métodos de Monte Carlo via Cadeia de Markov (MCMC) são alternativas aos métodos não interativos em problemas complexos. A ideia é simular uma cadeia de Markov irreduzível aperiódica cuja distribuição estacionária é a distribuição de interesse $\pi(\boldsymbol{\gamma}|y)$. Os

algoritmos de Metropolis-Hastings e *Gibbs Sampling* são os dois métodos de gerar amostras mais utilizados e são descritos na próxima subseção.

Gibbs Sampling

Esta técnica é usada para gerar variáveis aleatórias de uma distribuição sem utilizar diretamente sua densidade de probabilidade. O método consiste em um esquema iterativo de amostragem de uma cadeia de Markov cujo núcleo de transição é formado pelas distribuições condicionais completas.

Considere o vetor aleatório $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)'$, com distribuições condicionais completas a posteriori $\pi_i(\gamma_i) = \pi(\gamma_i | \boldsymbol{\gamma}_{-i})$, $i = 1, \dots, k$ estão disponíveis. Em muitas situações, a geração de uma amostra pode ser custosa, complicada ou simplesmente impossível. No entanto, se as condicionais apresentam formas conhecidas (normal, gama, beta, etc.), então o método de Gibbs fornece uma alternativa de geração baseada em sucessivas gerações das distribuições condicionais completas. O algoritmo é descrito da seguinte forma:

Passo 1: Arbitre os valores iniciais $\boldsymbol{\gamma}^{(0)} = (\gamma_1^{(0)}, \dots, \gamma_k^{(0)})$ e inicialize o contador de iterações da cadeia $j = 1$;

Passo 2: Obtenha um novo valor $\boldsymbol{\gamma}^{(j)} = (\gamma_1^{(j)}, \dots, \gamma_k^{(j)})$ através de sucessivas gerações de valores

$$\begin{aligned} \gamma_1^{(j)} &\sim \pi\left(\gamma_1 | \gamma_2^{(j-1)}, \gamma_3^{(j-1)}, \dots, \gamma_k^{(j-1)}\right) \\ \gamma_2^{(j)} &\sim \pi\left(\gamma_2 | \gamma_1^{(j)}, \gamma_3^{(j-1)}, \dots, \gamma_k^{(j-1)}\right) \\ &\vdots \\ \gamma_k^{(j)} &\sim \pi\left(\gamma_k | \gamma_1^{(j)}, \gamma_2^{(j)}, \dots, \gamma_{k-1}^{(j)}\right); \end{aligned}$$

Passo 3: Incrementar o contador de j para $j+1$ e retorne ao Passo 2 até obter convergência.

Cada iteração se completa a cada k movimentos, sendo k o número de parâmetros, ao longo dos eixos coordenados das componentes de $\boldsymbol{\gamma}$. À medida que o número de iterações aumenta, a cadeia se aproxima da sua condição de equilíbrio. Após a convergência, os valores resultantes formam uma amostra de $\pi(\boldsymbol{\gamma})$.

Metropolis-Hastings

O algoritmo de Metropolis-Hastings (Chib & Greenberg, 1995) é usado quando não são facilmente identificadas as distribuições condicionais a posteriori, o que impossibilita a geração direta a partir destas distribuições.

A ideia do método é gerar um valor de uma distribuição auxiliar e aceitar este valor com uma certa probabilidade. Este mecanismo de correção garante a convergência da cadeia para a distribuição de equilíbrio, neste caso a distribuição a posteriori.

Suponha que a cadeia esteja no estado γ e um valor γ^* é gerado de uma distribuição proposta $q(\cdot|\gamma)$. A distribuição proposta pode depender do estado atual da cadeia, por exemplo $q(\cdot|\gamma)$ pode ser uma distribuição normal centrada em γ . O novo valor γ^* é aceito com probabilidade

$$\alpha(\gamma, \gamma^*) = \min \left\{ 1, \frac{\pi(\gamma^*)q(\gamma|\gamma^*)}{\pi(\gamma)q(\gamma^*|\gamma)} \right\}, \quad (4.12)$$

sendo π a distribuição de interesse. A simulação é esquematizada da seguinte forma:

Passo 1: Arbitre o valor inicial $\gamma^{(0)}$ e inicialize o contador de iterações $j = 1$;

Passo 2: Gere um novo valor γ^* da distribuição $q(\cdot|\gamma)$;

Passo 3: Calcule a probabilidade de aceitação $\alpha(\gamma, \gamma^*)$ e gere $u \sim U(0, 1)$;

Passo 4: Se $u \leq \alpha$ aceite o novo valor e faça $\gamma^{(j+1)} = \gamma^*$, caso contrário rejeite e faça $\gamma^{(j+1)} = \gamma$;

Passo 5: Incrementar o contador de j para $j + 1$ e retorne ao Passo 2.

O conhecimento da densidade de probabilidade $\pi(\gamma)$ pode ser parcial, isto é, a menos de uma constante a probabilidade em (4.12) não se altera. A cadeia pode permanecer no mesmo estado por muitas iterações e na prática é comum monitorar a porcentagem das iterações para as quais novos valores são aceitos.

Na prática recomenda-se que as iterações iniciais sejam descartadas para evitar a influência das condições iniciais arbitrárias nas estimativas. Este procedimento é chamado de *burn-in*, ou período de aquecimento. Se a cadeia de Markov satisfaz certas condições

de regularidade, então quando o número de iterações aumenta, a cadeia converge para uma distribuição de equilíbrio, a qual é definida pela densidade a posteriori.

Existem vários métodos de verificação da convergência que são baseados nas propriedades da cadeia de Markov, para indicar se houve ou não a convergência da amostra simulada. Para verificar a convergência da cadeia, pode-se utilizar técnicas gráficas, tais como: gráficos de séries temporais, em que a série deve estar em uma faixa, sem apresentar tendências nem oscilações bruscas; gráficos de autocorrelação dos parâmetros, sendo esperado que a primeira correlação seja alta e as demais próximas de zero, indicando que a amostra simulada é não correlacionada. Uma outra avaliação de convergência é feita usando-se técnicas numéricas. Os diagnósticos de convergência mais utilizados são descritos por Gelman & Rubin (1992), em que consideram a análise de variância para verificar a convergência de cadeias paralelas (com valores iniciais super dispersos) e o critério Geweke (1992), que é baseado em métodos de séries temporais.

Neste trabalho, os métodos de MCMC apresentados são utilizados na obtenção de estimativas para quantidades de interesse do modelo. Também, para avaliar a convergência da cadeia, será utilizado o critério de Geweke, e este será denotado por z , conforme será visto nas tabelas de resumos a posteriori. Para isso, é apropriado que o valor da estatística de Geweke esteja no intervalo $-1.96 < z < 1.96$.

Para o caso em que as funções de transição são simétricas, pode-se simplificar o algoritmo de Metropolis-Hastings para o algoritmo de Metropolis, em que a única diferença é que a probabilidade de aceitação (4.12) será dada por

$$\alpha(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*) = \min \left\{ 1, \frac{\pi(\boldsymbol{\gamma}^*)}{\pi(\boldsymbol{\gamma})} \right\}.$$

4.4 Seleção de modelos

Assim como na metodologia clássica, a metodologia bayesiana também possui ferramentas para selecionar modelos ou então a ordem dos modelos. O próprio critério BIC (Seção 3.3) pode ser usado para tal feito, porém há critérios mais específicos e que se enquadram melhor nessa abordagem.

Raftery *et al.* (2006) propuseram uma outra versão para o BIC, chamada BICM ou BIC-Monte Carlo. Essa proposta tem como objetivo incrementar as estimativas de Monte Carlo de cada parâmetro no critério BIC de Schwarz (1978).

Dessa maneira, o BICM pode ser escrito como

$$BICM = \sum_{i=1}^M \{2\ell(\hat{\gamma}^{(i)}) + k \log(n)\}, \quad (4.13)$$

em que M é o tamanho do vetor de amostras selecionadas pelo processo de aceitação/rejeição do método MCMC em uso, $\ell(\hat{\gamma}^{(i)})$ é a log-verossimilhança avaliada em cada estimativa dos parâmetros, k é o número de parâmetros do modelo e n é o número de observações da série.

Seu uso é semelhante ao BIC usual, em que o modelo é escolhido de acordo com o valor de (4.13). O modelo que apresentar o menor BICM traz evidências de que esse seja o melhor.

4.5 Previsão

Da mesma maneira como na previsão para a inferência clássica (Seção 3.5), a previsão bayesiana também é feita utilizando o preditor linear, porém com algumas mudanças.

Ao longo da simulação MCMC, são geradas amostras para os parâmetros, as quais são rejeitadas ou não (Metropolis ou Metropolis-Hastings), ou então nunca são rejeitadas (Gibbs *sampling*). Neste trabalho, o método utilizado é o de Metropolis e, portanto, são guardadas amostras dos valores não rejeitados dos parâmetros, é dado um salto para poder retirar a correlação e, com os valores finais guardados, são calculadas a média, moda ou mediana para obter as estimativas dos parâmetros.

Dessa forma, não faz muito sentido calcular a previsão como na forma clássica, onde são utilizadas as estimativas dos parâmetros, neste caso a média, moda ou mediana. Na amostra final guardada, cada estimativa do parâmetro produzirá uma previsão diferente, sendo necessário, então, utilizar a média da previsão calculada para cada estimativa como a previsão obtida. Ou seja, as previsões 1-passo a frente para os três modelos são dadas

por

$$\hat{\mu}_{t+1} = \frac{1}{M} \sum_{i=1}^M \left[\hat{\alpha}_{0,i} + \sum_{j=1}^p \hat{\phi}_{j,i} y_{t+1-j} + \sum_{j=1}^q \hat{\theta}_{j,i} \{y_{t+1-j} - \mu_{t+1-j}\} \right]$$

para o caso do modelo Gaussiano,

$$\hat{\mu}_{t+1} = \frac{1}{M} \sum_{i=1}^M \left[\left(\hat{\alpha}_{0,i} + \sum_{j=1}^p \hat{\phi}_{j,i} \frac{1}{y_{t+1-j}^2} + \sum_{j=1}^q \hat{\theta}_{j,i} \left\{ \frac{1}{y_{t+1-j}^2} - \frac{1}{\mu_{t+1-j}^2} \right\} \right)^{-1/2} \right]$$

para o modelo Inversa Gaussiana e

$$\hat{\mu}_{t+1} = \frac{1}{M} \sum_{i=1}^M \left[\left(\hat{\alpha}_{0,i} + \sum_{j=1}^p \hat{\phi}_{j,i} \frac{1}{y_{t+1-j}} + \sum_{j=1}^q \hat{\theta}_{j,i} \left\{ \frac{1}{y_{t+1-j}} - \frac{1}{\mu_{t+1-j}} \right\} \right)^{-1} \right]$$

para o modelo Gama.

As previsões h -passos a frente são feitas analogamente à maneira clássica, porém com as médias tomadas como nas equações anteriores.

Capítulo 5

Aplicações

5.1 Aplicações com dados simulados

O uso de dados simulados, em geral, tem como objetivo verificar a viabilidade do modelo proposto, ou seja, a qualidade do ajuste e a qualidade das estimativas calculadas, já que se têm os verdadeiros valores dos parâmetros, entre outros. É possível também estudar propriedades de estimadores, propriedades assintóticas dos estimadores e a adequabilidade do modelo em relação ao conjunto de dados em questão.

Nesta seção é conduzido um estudo de simulação para o modelo GAR(2), com as distribuições Normal, Gama e Inversa Gaussiana. Após isso são feitos ajustes à algumas séries simuladas, com intuito de ilustrar o uso da abordagem bayesiana junto a abordagem clássica e uma aplicação com dados reais para ilustrar o uso do modelo. O algoritmo para a geração dos dados proposta neste trabalho encontra-se no Apêndice A.

Sobre o ajuste dos parâmetros de médias móveis

Conforme nota-se, não são utilizadas para os estudos aqui realizados a parte MA(q) dos modelos GARMMA. Na prática, o ajuste dos termos de médias móveis apresentam uma dificuldade técnica bastante grande, já que seu ajuste depende de médias que não são calculadas: as médias calculadas durante o processo correspondem às observações y_{r+1}, \dots, y_n (Seção 3.2), ou seja, $\mu_i, i = r + 1, \dots, n$.

Diante disso, os termos iniciais μ_1, \dots, μ_r devem ser aproximados e, em geral, não existem maneiras eficientes de calcular essa aproximação e isso traz à tona alguns problemas, dentre eles

- convergência não alcançada no método Escore de Fisher, e
- inversa da matriz dada na Equação 3.1 inexistente.

Os dois problemas acima foram os encontrados durante o processo de estimação realizado neste trabalho. Para ser mais preciso, durante as tentativas de geração de séries para o modelo GARMA Inversa Gaussiana, o método não obteve êxito no ajuste de nenhum valor prefixado dos parâmetros de médias móveis, bem como nenhuma ordem desses.

No caso dos modelos GARMA Normal e GARMA Gama, os únicos êxitos obtidos foram para valores prefixados bem pequenos dos parâmetros, digamos valores em torno de 0.1, 0.2 e 0.3. Valores negativos não foram estimados. Porém, apesar do método ter conseguido calcular as estimativas para aqueles valores, durante o estudo de simulação ainda haviam séries cujas inversas da matriz citada acima não existiam, fazendo com que os percentis de cobertura caíssem para algo em torno de 20%.

Para os dados reais, o problema ainda persistiu e não foi possível ajustar os dados com um modelo que incorporasse os parâmetros de médias móveis.

Na abordagem bayesiana, o erro ainda era frequente e fazia com que a taxa de aceitação/rejeição do Metropolis caísse para 0%, ou seja, todos os valores gerados eram rejeitados.

Em virtude desses problemas encontrados, as aplicações nesse trabalho contam apenas com os parâmetros autorregressivos do modelo.

5.1.1 Estudo de simulação

O estudo de simulação aqui apresentado tem como objetivo mostrar a consistência em média quadrática dos estimadores, definida a seguir.

Definição 5.1 *Para uma amostra de tamanho n de uma amostra Y , um estimador $\widehat{\gamma}$ é consistente em média quadrática se*

$$\lim_{n \rightarrow \infty} E \left[\left(\widehat{bm\gamma} - \gamma \right)^2 \right] = 0.$$

Além disso, para uma avaliação global do estudo, foram calculadas as médias das estimativas e dos desvios padrão médios. Foi calculado também o vício médio.

Foram então simulados os três modelos em estudo neste trabalho com suas respectivas funções de ligação canônicas. As observações foram geradas de uma componente linear com três parâmetros, ou seja, $\eta_t = \alpha_0 + \phi_1 g(y_{t-1}) + \phi_2 g(y_{t-2})$, que se trata de um modelo GAR(2), em que $g(\cdot)$ é a função de ligação. Os parâmetros de dispersão atribuídos foram os seguintes:

- $\varphi = 1$ e 0.1 para o modelo Gaussiano,
- $\varphi = 1$ e 0.5 para o modelo Inversa Gaussiana e
- $\varphi = 0.1$ e 0.05 para o modelo Gama.

O número de observações geradas (tamanho da amostra) foi $n = 30, 100$ e 1000 e o número de replicações foi de 10.000 .

Para cada combinação dos parâmetros que foram assumidos, foram ajustados os respectivos modelos e calculadas as estimativas de máxima verossimilhança de α_0, ϕ_1, ϕ_2 e φ . Intervalos de confiança assintóticos foram calculados para a obtenção do percentil de cobertura, bem como as médias dos parâmetros ajustados, média dos desvios padrões, vício e o erro quadrático médio foram calculados para observar o comportamento dos ajustes.

Os resultados dessas simulações para cada um dos modelos considerados, são apresentados nas Tabelas 5.1-5.6.

Tabela 5.1: Estudo de simulação para o modelo Normal com $\alpha_0 = 1$, $\phi_1 = 0.5$, $\phi_2 = -0.3$ e $\varphi = 1$.

n	EMV	% Cobertura	Média	D.P. Médio	Vício	EQM
30	$\hat{\alpha}_0$	95,30%	1.0958	0.3334	-0.0958	0.1229
	$\hat{\phi}_1$	94,96%	0.4564	0.1902	0.0435	0.0367
	$\hat{\phi}_2$	95,80%	-0.3328	0.1903	0.0328	0.0305
	$\hat{\varphi}$	82,49%	0.8969	0.2397	0.1030	0.0759
100	$\hat{\alpha}_0$	94,97%	1.0267	0.1703	-0.0267	0.0307
	$\hat{\phi}_1$	95,17%	0.4860	0.0977	0.0139	0.0097
	$\hat{\phi}_2$	95,44%	-0.3069	0.0978	0.0069	0.0091
	$\hat{\varphi}$	91,93%	0.9696	0.1386	0.0303	0.0207
1000	$\hat{\alpha}_0$	94,75%	1.0023	0.0524	-0.0023	0.0028
	$\hat{\phi}_1$	94,96%	0.4992	0.0301	0.0007	0.0009
	$\hat{\phi}_2$	94,95%	-0.3009	0.0301	0.0009	0.0009
	$\hat{\varphi}$	94,54%	0.9966	0.0446	0.0033	0.0019

Tabela 5.2: Estudo de simulação para o modelo Normal com $\alpha_0 = 1$, $\phi_1 = 0.5$, $\phi_2 = -0.3$ e $\varphi = 0.1$.

n	EMV	% Cobertura	Média	D.P. Médio	Vício	EQM
30	$\hat{\alpha}_0$	92,68%	1.0971	0.2560	-0.0971	0.0809
	$\hat{\phi}_1$	92,17%	0.4565	0.1748	0.0434	0.0370
	$\hat{\phi}_2$	93,18%	-0.3340	0.1748	0.0340	0.0312
	$\hat{\varphi}$	82,39%	0.0894	0.0239	0.0105	0.0007
100	$\hat{\alpha}_0$	94,10%	1.0251	0.1373	-0.0251	0.0199
	$\hat{\phi}_1$	94,15%	0.4865	0.0956	0.0134	0.0097
	$\hat{\phi}_2$	94,61%	-0.3069	0.0956	0.0069	0.0092
	$\hat{\varphi}$	90,73%	0.0968	0.0138	0.0031	0.0002
1000	$\hat{\alpha}_0$	94,98%	1.0023	0.0430	-0.0023	0.0018
	$\hat{\phi}_1$	94,97%	0.4988	0.0301	0.0011	0.0009
	$\hat{\phi}_2$	94,91%	-0.3005	0.0301	0.0005	0.0009
	$\hat{\varphi}$	94,17%	0.0996	0.0044	0.0003	2e-05

Tabela 5.3: Estudo de simulação para o modelo Inversa Gaussiana com $\alpha_0 = 1$, $\phi_1 = 0.5$, $\phi_2 = 0.1$ e $\varphi = 1$.

n	EMV	% Cobertura	Média	D.P. Médio	Vício	EQM
30	$\hat{\alpha}_0$	90,34%	4.5882	3.1817	-3.5882	51.371
	$\hat{\phi}_1$	89,33%	0.4449	0.1608	0.0550	0.0318
	$\hat{\phi}_2$	91,64%	0.0630	0.1270	0.0369	0.0168
	$\hat{\varphi}$	82,40%	0.8898	0.2378	0.1101	0.0771
100	$\hat{\alpha}_0$	93,88%	1.7728	1.1153	-0.7628	2.6736
	$\hat{\phi}_1$	93,64%	0.4850	0.0884	0.0149	0.0082
	$\hat{\phi}_2$	94,55%	0.0905	0.0688	0.0094	0.0045
	$\hat{\varphi}$	91,17%	0.9695	0.1385	0.0304	0.0210
1000	$\hat{\alpha}_0$	94,59%	1.0622	0.2916	-0.0622	0.0945
	$\hat{\phi}_1$	94,79%	0.4986	0.0281	0.0013	0.0008
	$\hat{\phi}_2$	94,85%	0.0992	0.0218	0.0008	0.0005
	$\hat{\varphi}$	94,67%	0.9970	0.0446	0.0030	0.0020

Tabela 5.4: Estudo de simulação para o modelo Inversa Gaussiana com $\alpha_0 = 1$, $\phi_1 = 0.5$, $\phi_2 = 0.1$ e $\varphi = 0.5$.

n	EMV	% Cobertura	Média	D.P. Médio	Vício	EQM
30	$\hat{\alpha}_0$	90,49%	2.4413	1.4727	-1.4413	7.52
	$\hat{\phi}_1$	89,64%	0.4462	0.1657	0.0537	0.0343
	$\hat{\phi}_2$	91,72%	0.0574	0.1378	0.0425	0.0202
	$\hat{\varphi}$	82,08%	0.4439	0.1186	0.0560	0.0190
100	$\hat{\alpha}_0$	94,01%	1.3378	0.6137	-0.3378	0.6041
	$\hat{\phi}_1$	93,71%	0.4850	0.0906	0.0149	0.0086
	$\hat{\phi}_2$	94,34%	0.0889	0.0745	0.0110	0.0054
	$\hat{\varphi}$	91,58%	0.4845	0.0692	0.0154	0.0051
1000	$\hat{\alpha}_0$	94,78%	1.0275	0.1744	-0.0275	0.0325
	$\hat{\phi}_1$	94,75%	0.4983	0.0287	0.0016	0.0008
	$\hat{\phi}_2$	94,91%	0.0991	0.0235	0.0008	0.0005
	$\hat{\varphi}$	95,09%	0.4984	0.0223	0.0015	0.0004

Tabela 5.5: Estudo de simulação para o modelo Gama com $\alpha_0 = 0.8$, $\phi_1 = 0.5$, $\phi_2 = 0.2$ e $\varphi = 0.1$.

n	EMV	% Cobertura	Média	D.P. Médio	Vício	EQM
30	$\hat{\alpha}_0$	87,94%	1.3881	0.3844	-0.5800	0.8741
	$\hat{\phi}_1$	89,62%	0.4321	0.0306	0.0678	0.0412
	$\hat{\phi}_2$	90,96%	0.1200	0.0282	0.0799	0.0349
	$\hat{\varphi}$	81,55%	0.0882	0.0232	0.0117	0.0007
100	$\hat{\alpha}_0$	92,90%	0.9539	0.0755	-0.1539	0.1092
	$\hat{\phi}_1$	93,53%	0.4837	0.0088	0.0162	0.0097
	$\hat{\phi}_2$	93,96%	0.1755	0.0080	0.0244	0.0085
	$\hat{\varphi}$	91,04%	0.0968	0.0136	0.0031	0.0002
1000	$\hat{\alpha}_0$	94,80%	0.8145	0.0795	-0.0145	0.0067
	$\hat{\phi}_1$	95,04%	0.4983	0.0295	0.0017	0.0008
	$\hat{\phi}_2$	95,05%	0.1977	0.0281	0.0022	0.0007
	$\hat{\varphi}$	94,65%	0.0996	0.0043	0.0003	1.9e-05

Tabela 5.6: Estudo de simulação para o modelo Gama com $\alpha_0 = 0.8$, $\phi_1 = 0.5$, $\phi_2 = 0.2$ e $\varphi = 0.05$.

n	EMV	% Cobertura	Média	D.P. Médio	Vício	EQM
30	$\hat{\alpha}_0$	88,69%	1.2972	0.5351	-0.4972	0.6535
	$\hat{\phi}_1$	89,63%	0.4298	0.1792	0.0701	0.0453
	$\hat{\phi}_2$	91,02%	0.1127	0.1757	0.0872	0.0388
	$\hat{\varphi}$	81,73%	0.0443	0.0117	0.0056	0.0001
100	$\hat{\alpha}_0$	92,81%	0.9311	0.2524	-0.1311	0.0903
	$\hat{\phi}_1$	93,44%	0.4843	0.0963	0.0156	0.0102
	$\hat{\phi}_2$	93,80%	0.1738	0.0940	0.0261	0.0096
	$\hat{\varphi}$	90,81%	0.0484	0.0068	0.0015	5.2e-05
1000	$\hat{\alpha}_0$	95,06%	0.8124	0.0753	-0.0124	0.0058
	$\hat{\phi}_1$	94,80%	0.4982	0.0302	0.0017	0.0009
	$\hat{\phi}_2$	95,06%	0.1976	0.0294	0.0023	0.0008
	$\hat{\varphi}$	94,84%	0.0498	0.0022	0.0001	4.9e-06

As Figuras 5.1 e 5.2 têm objetivo de ilustrar o comportamento do percentil de cobertura em função do tamanho da amostra.

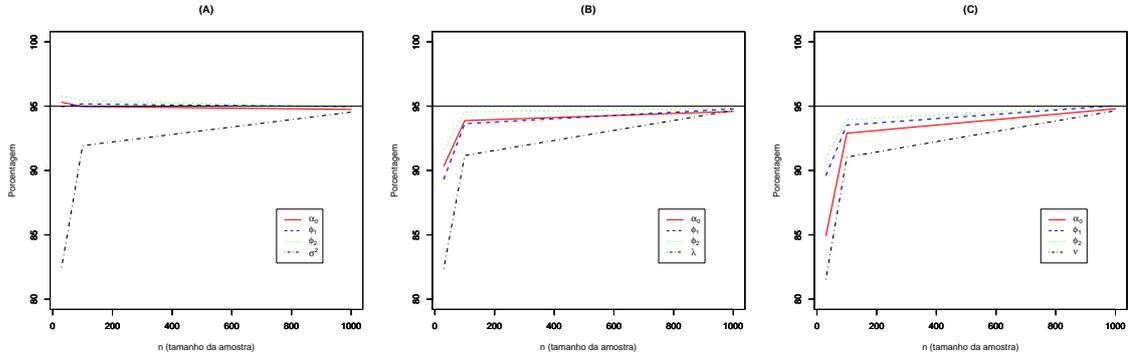


Figura 5.1: Percentil de cobertura para os modelos (A) gaussiano; (B) inverso gaussiano, e (C) gama, com $\varphi = 1, 1$ e 0.1 respectivamente.

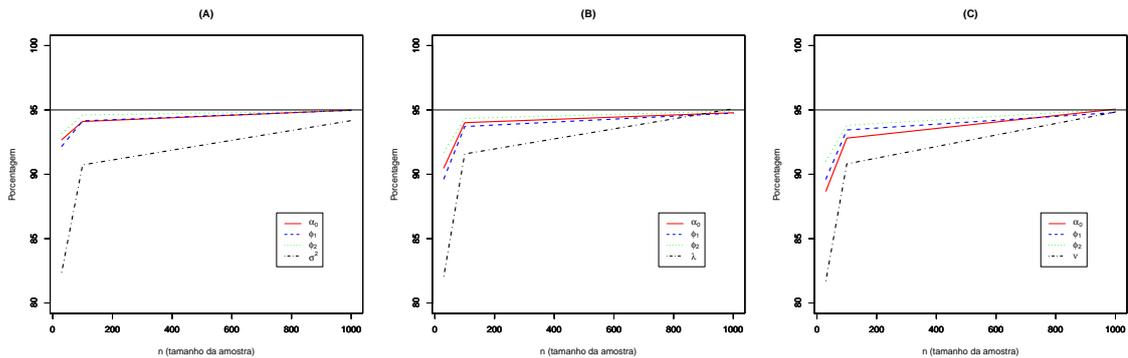


Figura 5.2: Percentil de cobertura para os modelos (A) gaussiano; (B) inverso gaussiano, e (C) gama, com $\varphi = 0.1, 0.5$ e 0.05 respectivamente.

Também pode-se avaliar o comportamento da amplitude do intervalo de confiança ao longo das replicações, fazendo uma média das amplitudes de cada intervalo de confiança. As Figuras 5.3 e 5.4 podem ilustrar como a amplitude tende a diminuir conforme o aumento do tamanho amostral.

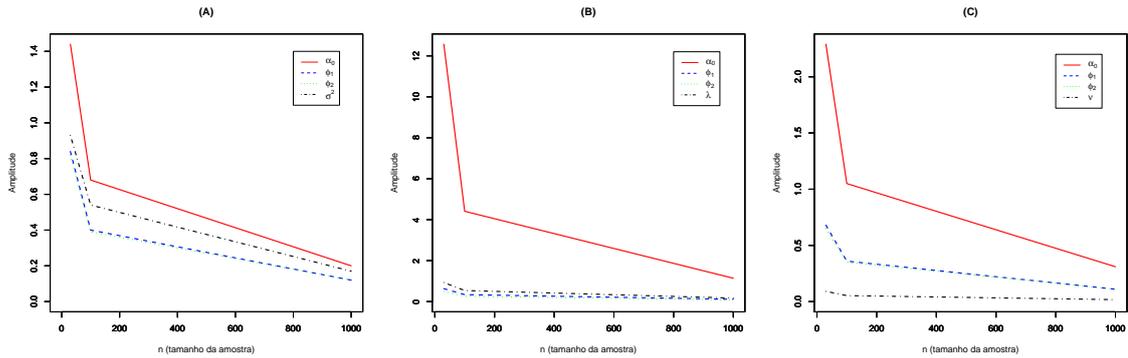


Figura 5.3: Amplitudes médias para os modelos (A) gaussiano; (B) inverso gaussiano, e (C) gama, com $\varphi = 1, 1$ e 0.1 respectivamente.

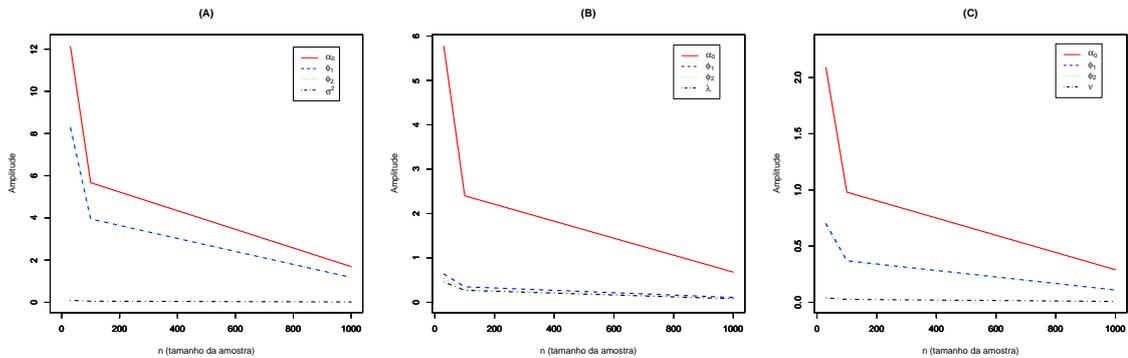


Figura 5.4: Amplitudes médias para os modelos (A) gaussiano; (B) inverso gaussiano, e (C) gama, com $\varphi = 0.1, 0.5$ e 0.05 respectivamente.

Conforme pode ser observado nos resultados apresentados nas Tabelas 5.1 à 5.6 e nos gráficos das Figuras 5.1 à 5.4, verifica-se a consistência dos estimadores conforme aumenta-se o tamanho da amostra mesmo variando-se o parâmetro de dispersão.

As propriedades assintóticas são verificadas com o percentil de cobertura aproximando-se dos 95%, bem como os valores dos parâmetros estimados tornaram-se mais próximos dos valores reais. Nota-se também que o vício e o EQM dos estimadores aproximam-se de zero indicando que para os modelos simulados há indícios de que os estimadores são consistentes e não viciados.

5.1.2 Ajustes clássicos e bayesianos para séries simuladas

O objetivo desta seção está no fato de avaliar se as abordagens clássica e bayesiana são eficazes no ajuste de séries com dados simulados. Uma vez que os verdadeiros valores dos parâmetros são conhecidos, pode-se ter uma real ideia se os métodos tem um bom desempenho relacionado ao ajuste. Para isso, foram geradas novamente observações das distribuições normal, gama e inversa gaussiana.

Na abordagem bayesiana desse problema foi gerada da densidade a posteriori, uma amostra total de tamanho 50000 para cada parâmetro de cada modelo. Desprezou-se 2000 observações como período de aquecimento e da amostra restante foram selecionadas um valor a cada 10 gerados para diminuir o número de repetições devido ao processo de aceitação/rejeição. Esse procedimento resultou em uma amostra de 4800 observações para cada parâmetro. Foi feito um controle da taxa de aceitação para que a mesma pudesse ficar em torno de 30%.

Um sumário da densidade a posteriori obtidas pelo método de Monte Carlo são apresentados nas Tabelas 5.7, 5.9 e 5.11 para o modelo GAR(2), para cada uma das distribuições consideradas. Para fins de referência, são apresentados nas Tabelas 5.8, 5.10 e 5.12 as estimativas de máxima verossimilhança para cada um dos modelos considerados.

Modelo GAR Normal

Para o modelo GAR Normal, foi gerada uma série de tamanho $n = 400$, com $\alpha_0 = 1$, $\phi_1 = 0.5$, $\phi_2 = -0.3$ e $\varphi = 1$. O resumo das distribuições marginais a posteriori e os EMVs são dados nas tabelas a seguir.

Tabela 5.7: Resumo das distribuições marginais *a posteriori* para o modelo normal.

	Valor Real	Média	Mediana	Moda	D.P.	<i>I.Cred. de 95%</i>
$\hat{\alpha}_0$	1	1.0630	1.0640	1.0778	0.0701	[0.9254 ; 1.1992]
$\hat{\phi}_1$	0.5	0.4579	0.4575	0.4609	0.0387	[0.3817 ; 0.5350]
$\hat{\phi}_2$	-0.3	-0.3024	-0.3026	-0.3027	0.0389	[-0.3798 ; -0.2271]
$\hat{\varphi}$	1	0.9923	0.9908	0.9828	0.0494	[0.8985 ; 1.0917]

Tabela 5.8: Estimativas de Máxima Verossimilhança para o modelo normal.

	Estimativas	D.P.	I.Conf. de 95%
$\hat{\alpha}_0$	1.0647	0.0844	[0.8984 ; 1.2309]
$\hat{\phi}_1$	0.4571	0.0475	[0.2908 ; 0.6233]
$\hat{\phi}_2$	-0.3024	0.0474	[-0.4686 ; -0.1361]
$\hat{\varphi}$	0.9904	0.0702	[0.8528 ; 1.1281]

Comparando-se os resultados apresentados nas Tabelas 5.7 e 5.8, observamos valores muito próximos das estimativas de máxima verossimilhança e da média a posteriori. Nos gráficos das Figuras 5.5 são apresentadas a série gerada e a série ajustada, e o gráfico de dispersão dos valores gerados e da média ajustada.

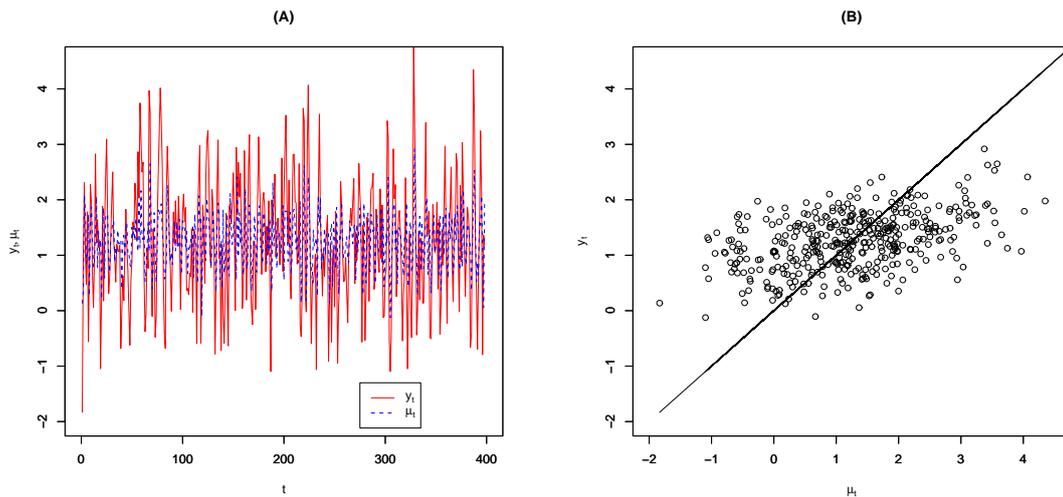


Figura 5.5: (A) Série simulada e série ajustada e (B) Dispersão dos dados ajustados com o modelo normal.

Note que o ajuste da série pela abordagem bayesiana apresenta-se bastante satisfatório e traz evidências de que se trata de um bom método para o ajuste.

Modelo GAR Gama

Da mesma forma que no modelo GAR Normal, para o modelo GAR Gama foi gerada uma série de tamanho $n = 300$, com $\alpha_0 = 0.8$, $\phi_1 = 0.5$, $\phi_2 = 0.2$ e $\varphi = 0.1$. O resumo das distribuições marginais a posteriori e os EMVs são dados nas tabelas a seguir.

Tabela 5.9: Resumo das distribuições marginais *a posteriori* para o modelo gama.

	Valor Real	Média	Mediana	Moda	D.P.	<i>I.Cred. de 95%</i>
$\hat{\alpha}_0$	0.8	0.7543	0.7550	0.7609	0.1094	[0.5396 ; 0.9625]
$\hat{\phi}_1$	0.5	0.5405	0.5398	0.5362	0.0461	[0.4521 ; 0.6345]
$\hat{\phi}_2$	0.2	0.1717	0.1719	0.1735	0.0435	[0.0843 ; 0.2583]
$\hat{\phi}$	0.1	0.0924	0.0924	0.0924	0.0007	[0.0910 ; 0.0937]

Tabela 5.10: Estimativas de Máxima Verossimilhança para o modelo gama.

	Estimativas	D.P.	<i>I.Conf. de 95%</i>
$\hat{\alpha}_0$	0.7492	0.1039	[0.4994 ; 0.9991]
$\hat{\phi}_1$	0.5410	0.0547	[0.4350 ; 0.6470]
$\hat{\phi}_2$	0.1721	0.0501	[0.0726 ; 0.2716]
$\hat{\phi}$	0.0923	0.0074	[0.0777 ; 0.1069]

Pelos valores apresentados nas Tabelas 5.9 e 5.10, não se percebe uma diferença marcante nas estimativas de máxima verossimilhança e na média a posteriori, o que indica que ambos ajustes estão bem próximos. Na Figura 5.6 aparecem o gráfico do ajuste e da dispersão, respectivamente.

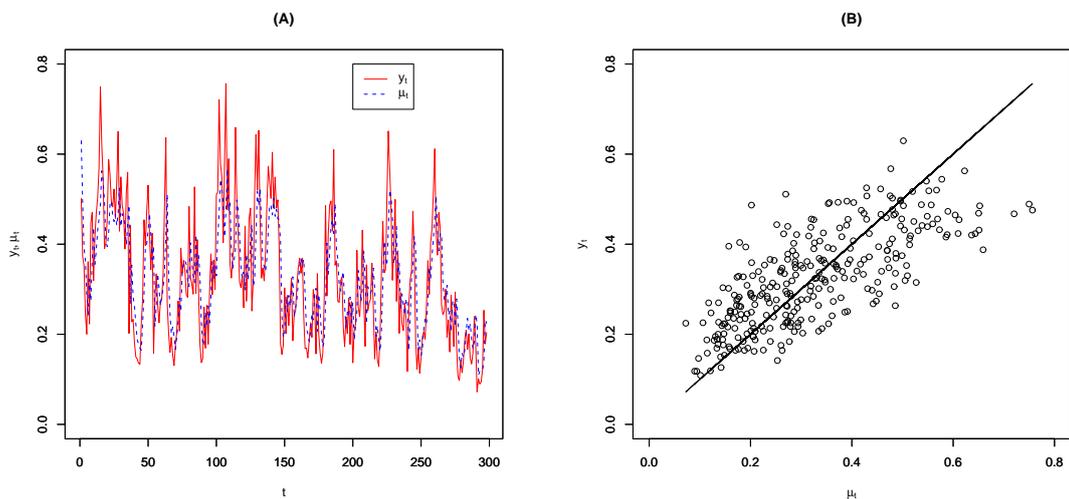


Figura 5.6: (A) Série simulada e série ajustada e (B) Dispersão dos dados ajustados com o modelo gama.

Para o modelo GAR Gama a abordagem bayesiana também parece ser uma boa opção para a modelagem da série.

Modelo GAR Inverso Gaussiano

Por fim, para o modelo GAR Inverso Gaussiano foi gerada uma série também de tamanho $n = 300$, com $\alpha_0 = 8$, $\phi_1 = 0.5$, $\phi_2 = 0.1$ e $\varphi = 1$. O resumo das distribuições marginais a posteriori e os EMVs são dados nas tabelas a seguir.

Tabela 5.11: Resumo das distribuições marginais *a posteriori* para o modelo inverso gaussiano.

	Valor Real	Média	Mediana	Moda	D.P.	<i>I.Cred. de 95%</i>
$\hat{\alpha}_0$	8	10.303	10.332	10.947	2.4924	[5.4839 ; 15.245]
$\hat{\phi}_1$	0.5	0.5486	0.5492	0.5522	0.0436	[0.4648 ; 0.6357]
$\hat{\phi}_2$	0.1	0.1569	0.1569	0.1541	0.0392	[0.0777 ; 0.2349]
$\hat{\varphi}$	1	0.9997	0.9975	0.9961	0.0571	[0.8914 ; 1.1152]

Tabela 5.12: Estimativas de Máxima Verossimilhança para o modelo inverso gaussiano.

	Estimativas	D.P.	<i>I.Conf. de 95%</i>
$\hat{\alpha}_0$	10.225	3.1263	[4.0973 ; 16.3527]
$\hat{\phi}_1$	0.5480	0.0542	[0.4418 ; 0.6543]
$\hat{\phi}_2$	0.1568	0.0486	[0.0614 ; 0.2522]
$\hat{\varphi}$	0.9940	0.0814	[0.8344 ; 1.1536]

Como nos modelos Normal e Gama, fica claro que também não há uma diferença significativa das estimativas de ambos os métodos pelas Tabelas 5.11 e 5.12. A Figura 5.7 mostra o ajuste e a dispersão, respectivamente.

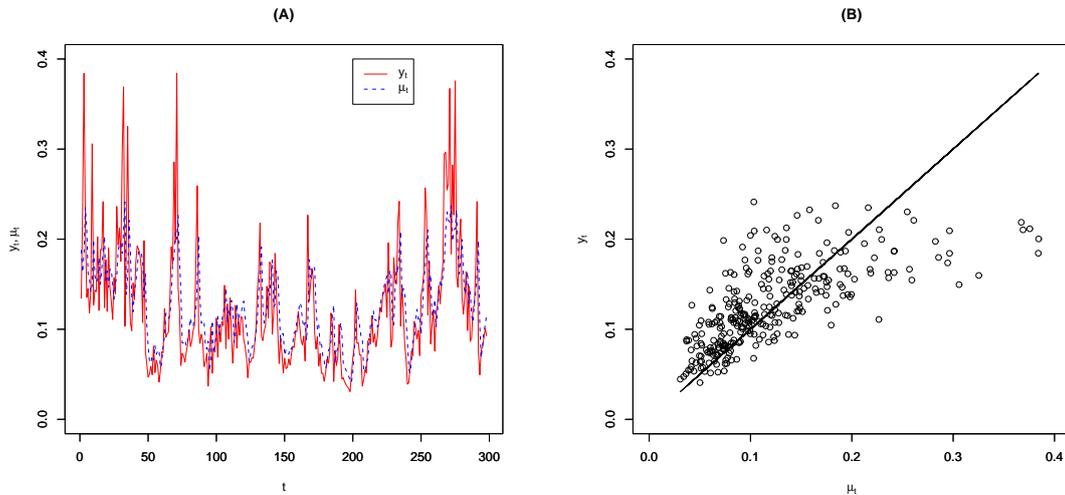


Figura 5.7: (A) Série simulada e série ajustada e (B) dispersão dos dados ajustados com o modelo inverso gaussiano.

Fica claro pela análise gráfica que a abordagem bayesiana para o modelo GAR Inversa Gaussiana também é bastante satisfatória.

Os valores do critério de Geweke (1992), utilizado para verificar a convergência de cada parâmetro gerado, foram:

- $z_{\hat{\alpha}_0} = -0.77$, $z_{\hat{\phi}_1} = -0.63$, $z_{\hat{\phi}_2} = 0.35$ e $z_{\hat{\varphi}} = -0.36$, para o modelo normal;
- $z_{\hat{\alpha}_0} = -0.13$, $z_{\hat{\phi}_1} = -0.19$, $z_{\hat{\phi}_2} = 0.17$ e $z_{\hat{\varphi}} = 0.67$, para o modelo gama, e
- $z_{\hat{\alpha}_0} = -0.26$, $z_{\hat{\phi}_1} = -0.79$, $z_{\hat{\phi}_2} = 0.71$ e $z_{\hat{\varphi}} = 1.62$ para o modelo inverso gaussiano.

De acordo com as estimativas, nota-se que há indícios de que houve convergência nas simulações, já que esta é alcançada quando $|z_j| < 1.96$.

Pode-se perceber, de acordo com os resultados obtidos acima com uso de dados simulados, que as estimativas de ambas abordagens ficaram bem próximas dos valores conhecidos dos parâmetros, evidenciando de que as duas abordagens são boas em ajustar esse tipo de modelo.

5.2 Aplicação com dados reais

Nesta seção é feita uma aplicação com a utilização de um conjunto de dados reais. O propósito geral é mostrar a aplicabilidade e a precisão dos modelos GARMA na modelagem de séries temporais reais.

5.2.1 Taxa de casos de tuberculose na região metropolitana de SP.

Aqui são propostos alguns submodelos do modelo GARMA para modelar uma série da taxa de casos de tuberculose a cada 100 mil pessoas na região metropolitana de São Paulo.

A tuberculose pulmonar é uma infecção causada por um microorganismo conhecido como Bacilo de Koch. Além dos pulmões, também pode ocorrer em outros órgãos do corpo, mesmo sem causar algum dano pulmonar. Somente no ano de 1990, a OMS estimou a presença de 8 milhões de novos casos, sendo que desses, 2.6 milhões de pessoas morreram. Especificamente no Brasil, em 1996 ocorreram 5928 mortes atribuídas a tuberculose e em 1998, foi presenciado cerca de 51,3 casos para cada 100 mil habitantes, sendo grande parte encontrada em regiões socioeconomicamente desfavorecidas, como norte e nordeste.

Uma das formas de contágio da tuberculose é através do ar contaminado eliminado pelo indivíduo que contraiu a doença. Se o indivíduo que inalar gotículas de saliva eliminada pela tosse ou espirro de um indivíduo contaminado, ele pode contrair a doença, sendo que esta pode se propagar se o sistema de defesa estiver em más condições, e por este motivo ela é encontrada em regiões menos favorecidas. Caso contrário o sistema de defesa elimina o bacilo.

Um grande vilão para o sistema imune é o vírus da Síndrome da Imunodeficiência Humana (HIV), pois este degenera o sistema imune e propicia o desenvolvimento da tuberculose nas pessoas portadoras. Para se ter base de comparação, a chance do indivíduo infectado pelo HIV de contrair tuberculose é de aproximadamente 10% ao ano, enquanto que em um indivíduo sadio, esta porcentagem é ao longo da vida. Por isso, com o surgimento do HIV, houve um crescimento muito grande de pessoas infectadas pela tuberculose. Para maiores informações a respeito, ver um excelente texto no site

www.abcdasaude.com.br/artigo.php?432.

Este conjunto de dados corresponde a uma série mensal de internações por tuberculose na cidade de São Paulo durante o período de janeiro de 1998 a setembro de 2009. Para as séries de internações consideramos a taxa de internações para cada 100.000 habitantes obtidas no Sistema de Informações Hospitalares do Ministério da Saúde. No Código Internacional de Doenças (CID10), a tuberculose é identificada pelos códigos A15 a A19 (Capítulo X). Possui um total de 141 observações com uma tendência decrescente, conforme mostra a Figura 5.8.

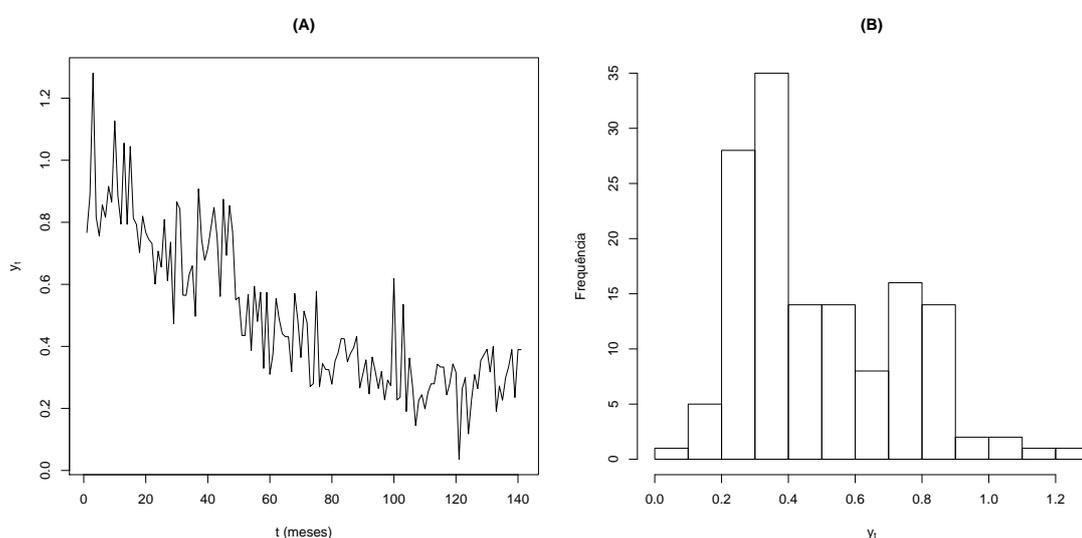


Figura 5.8: (A) Série da taxa de casos de tuberculose na região metropolitana de São Paulo; (B) Histograma dos dados.

De acordo os gráficos apresentados na Figura 5.8, nota-se uma leve assimetria no histograma dos dados. Isso pode ser um indício de que o modelo Normal não seja bom o bastante para a modelagem, sugerindo o uso de uma transformação ou o uso de um modelo que incorpore em sua distribuição a assimetria, como é o caso da distribuição Gama.

5.2.2 Método Escore de Fisher

A análise foi realizada com três submodelos do GAR, o modelo normal, gama e normal com transformação. A transformação utilizada foi a de Box & Cox (1964) (ver Apêndice B) com o parâmetro $\lambda = 0.5$. O uso da transformação de Box & Cox (1964) tem por

objetivo principal normalizar os dados, e foi utilizada aqui para verificar o comportamento do modelo GARMA com o uso dessa classe de transformações.

A ordem dos modelos ajustado variou de 1 a 4, sendo que os resultados obtidos são descritos na Tabela 5.13.

Tabela 5.13: Estimativas de Máxima Verossimilhança para o modelo GAR(p).

GAR(p)	Normal		Gama		Normal Transformada		
	Estimativa	D.P.	Estimativa	D.P.	Estimativa	D.P.	
$p = 1$	α_0	0.1041	0.0286	0.1366	0.5111	-0.1509	0.0384
	ϕ_1	0.7854	0.0517	0.0636	0.4388	0.7663	0.0537
	φ	0.0218	0.0026	0.0166	0.1096	0.0466	0.0055
$p = 2$	α_0	0.0556	0.0263	0.6137	0.1286	-0.0907	0.0358
	ϕ_1	0.4060	0.0741	0.3279	0.0707	0.4022	0.0746
	ϕ_2	0.4718	0.0738	0.3130	0.0703	0.4642	0.0743
	φ	0.0167	0.0020	0.1208	0.0142	0.0364	0.0043
$p = 3$	α_0	0.0419	0.0233	0.4297	0.1069	-0.0766	0.0320
	ϕ_1	0.1951	0.0736	0.1338	0.0506	0.1809	0.0743
	ϕ_2	0.3228	0.0709	0.1323	0.0500	0.2879	0.0719
	ϕ_3	0.3757	0.0733	0.4845	0.0683	0.4314	0.0742
	φ	0.0126	0.0015	0.0823	0.0097	0.0278	0.0033
$p = 4$	α_0	0.0350	0.0233	0.4099	0.1056	-0.0727	0.0323
	ϕ_1	0.1341	0.0839	0.0872	0.0472	0.1107	0.0843
	ϕ_2	0.2659	0.0744	0.1095	0.0473	0.2405	0.0751
	ϕ_3	0.3321	0.0754	0.4330	0.0679	0.3977	0.0758
	ϕ_4	0.1729	0.0788	0.1308	0.0566	0.1601	0.0820
	φ	0.0122	0.0014	0.0788	0.0094	0.0273	0.0033

Os modelos com distribuição Normal podem não ser aconselháveis porque não incorporam assimetria em sua distribuição, que é fortemente recomendada para esse conjunto de dados. Também ocorre que algumas estimativas de parâmetros do modelo foram não significativas, pois o intervalo de confiança cobre o valor zero. Portanto o modelo normal não será utilizado para dar continuidade na modelagem da série.

Já o modelo Normal com transformação, apesar de normalizar os dados, possui o inconveniente de toda a análise ser feita com base na série transformada. Isto acarreta em algumas dificuldades nos cálculos das previsões, já que será necessário fazer uso de métodos numéricos para inverter a transformação e obter a previsão da série na escala original. Além disso, neste ajuste foi escolhido arbitrariamente um parâmetro de transformação e este pode não ser o parâmetro mais indicado. Seria então necessário encontrar o parâmetro ótimo para a transformação através do cálculo da função de verossimilhança perfilada, o que tornaria esta análise muito mais complexa.

Finalmente tem-se o modelo gama que, além de não necessitar de transformação, é uma distribuição assimétrica como já foi dito. Ao contrário do modelo normal, não possui nenhuma de suas estimativas dos parâmetros não significativos nas ordens ajustadas. Isso torna a análise mais simples e, por esse motivo, o modelo gama será escolhido para ajustar o modelo.

Para a escolha da ordem do modelo foi utilizada a métrica BIC (Seção 3.3) e escolhe-se o modelo mais adequado para o ajuste aos dados o modelo que apresentar o menor valor do BIC.

Tabela 5.14: BIC para o modelo GAR Gama(p).

Ordem	GAR(1)	GAR(2)	GAR(3)	GAR(4)
BIC	-92.73	-100.95	-122.01	-119.71

De acordo com a Tabela 5.14, o modelo escolhido é o modelo GAR Gama(3). Na Tabela 5.15, segue as estimativas para os parâmetros.

Tabela 5.15: Estimativas de Máxima Verossimilhança.

	Estimativas	D.P.	<i>I.Conf. de 95%</i>
$\hat{\alpha}_0$	0.4297	0.1069	[0.2202 ; 0.6393]
$\hat{\phi}_1$	0.1338	0.0506	[0.0345 ; 0.2332]
$\hat{\phi}_2$	0.1323	0.0500	[0.0342 ; 0.2305]
$\hat{\phi}_3$	0.4845	0.0683	[0.3506 ; 0.6184]
$\hat{\phi}$	0.0823	0.0097	[0.0631 ; 0.1014]

Na Figura 5.9, seguem os gráficos do ajuste e de dispersão, respectivamente.

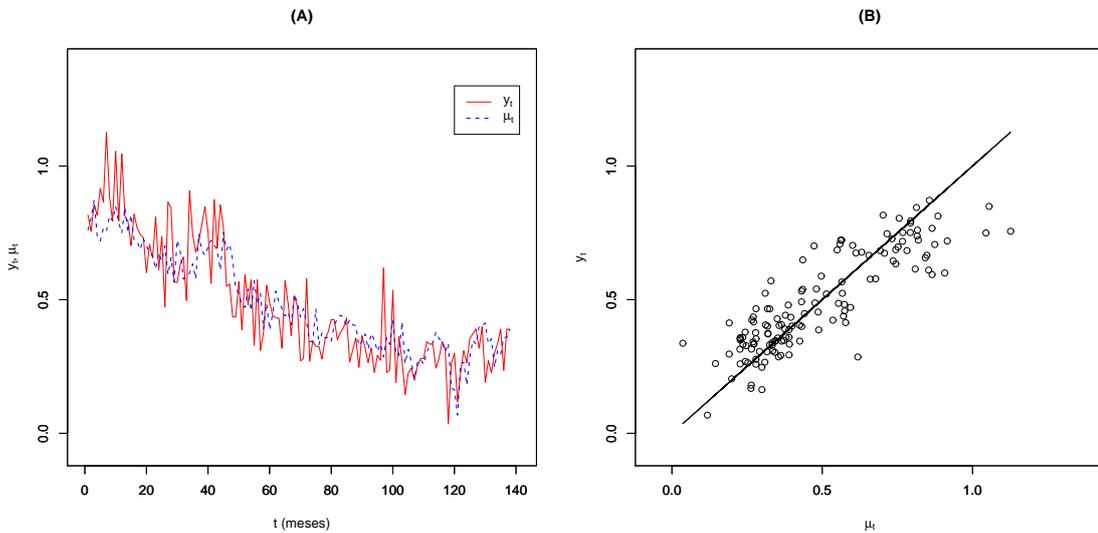


Figura 5.9: (A) Série e média ajustada; (B) Dispersão dos dados ajustados.

Também foram calculados o erro quadrático médio (MSE) e o erro médio absoluto percentual (MAPE), dados em (3.5) e (3.6) respectivamente. Os resultados obtidos foram $MSE = 25.61\%$ e $MAPE = 26.37\%$.

Através dos resultados obtidos acima, pode-se concluir que o modelo GAR Gama(3) mostra-se bastante eficiente em estimar a média da série, mesmo esta tendo uma tendência decrescente, em que um modelo ARMA não seria o mais indicado.

Análise dos Resíduos

Conforme escrito na Seção 3.4, o próximo passo da modelagem é o diagnóstico do modelo, que é feita através da análise dos resíduos do ajuste.

Como não se pode garantir a normalidade para os resíduos, a análise será feita com o histograma e as funções de autocorrelação e autocorrelação parcial. Também faz-se necessário a análise do gráfico de dispersão para verificar se há a presença de tendências, o que pode indicar variância não constante para os resíduos.

A Figura 5.10 traz os gráficos para a análise.

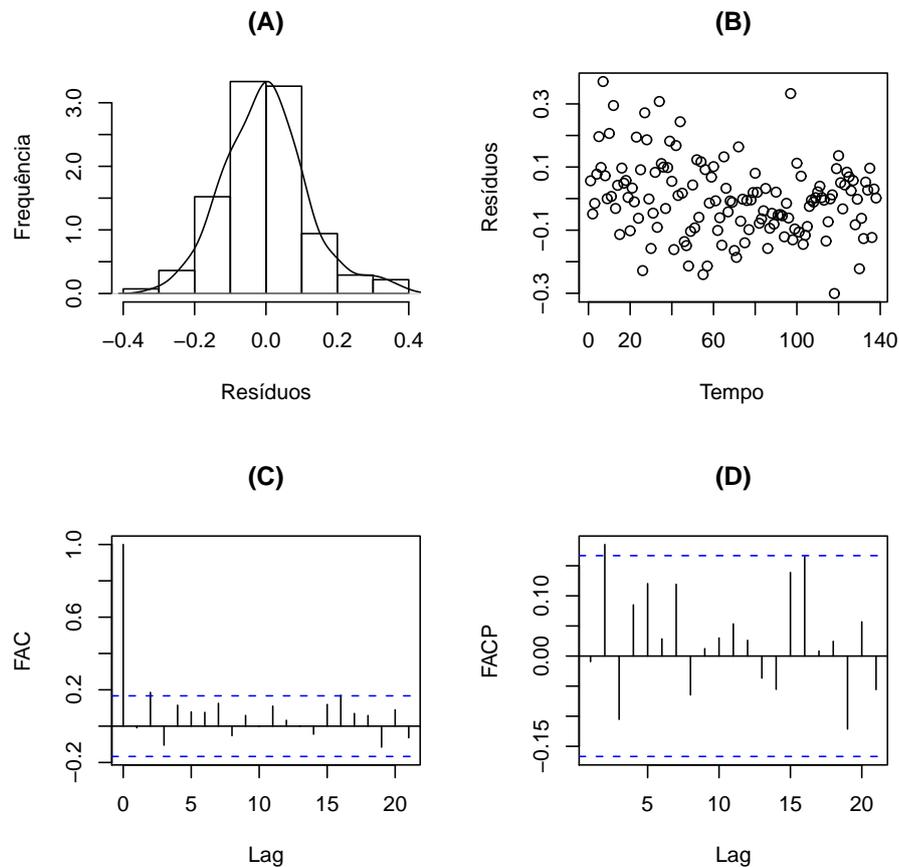


Figura 5.10: (A) Histograma; (B) Dispersão, (C) Função de autocorrelação e (D) Função de autocorrelação parcial.

De acordo com os gráficos apresentados na Figura 5.10, pode-se dizer que os resíduos apresentados possuem média 0 e variância constante, uma vez que o gráfico de dispersão não apresenta nenhuma tendência aparente. Além do mais, os gráficos da FAC e FACP dão indícios de que os resíduos são não correlacionados, o que pode ser comprovado de acordo com a estatística de Box-Ljung, cujo valor é de 15.37 com base em lag de tamanho 15, e a estatística $\chi_{12,0.05}^2 = 21.02$.

Com base nas análises feitas anteriormente, pode-se concluir que o modelo escolhido fornece um ajuste adequado ao conjunto de dados.

Previsão

A previsão para o conjunto de dados aqui apresentado será feita conforme descrito na Seção 3.5, utilizando o modelo GAR $\text{gama}(3)$, que foi o modelo que melhor se ajusta aos dados. O tamanho do passo de previsão será de $h = 5$, sendo que são previstas as 5 últimas observações da série que foram retiradas do conjunto e será utilizado o Erro Percentil Absoluto Médio (EPAM) dado em (3.6), para verificar a qualidade da previsão.

Tabela 5.16: Previsão para $h = 5$ meses

Nº da observação	Valor original	Valor previsto
137	0.3354	0.3540
138	0.3897	0.3536
139	0.2355	0.3948
140	0.3895	0.3981
141	0.3895	0.4046

De acordo com a Tabela 5.16, pode-ser perceber que os valores previstos tendem a manter uma média, sem a variância que os valores retirados possuem, o que é de se esperar dado que os valores calculados são os $\hat{\mu}$ dados no preditor linear. O EPAM calculado para avaliar a previsão, ficou em torno de 17%, evidenciando que a previsão feita apresenta valores coerentes. A figura 5.11 mostra os valores observados e previstos.

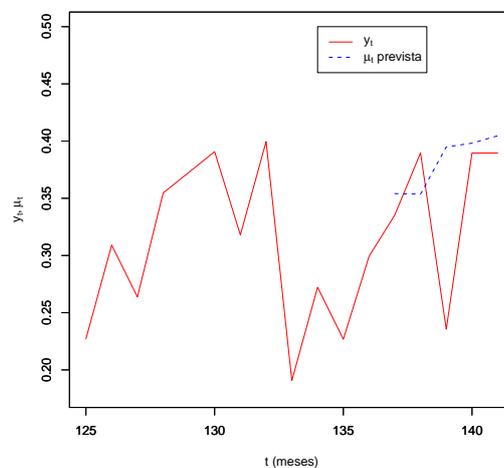


Figura 5.11: Previsão 5 meses a frente.

Nota-se também que a leve tendência crescente existente na parte final da série é captada e que, através dessas conclusões, o modelo mostra ter uma capacidade de previsão eficiente.

5.2.3 Abordagem bayesiana

O mesmo conjunto de dados, agora, será modelado via abordagem bayesiana. As distribuições a priori utilizadas foram não-informativas, ou seja, não é dada qualquer tipo de informação ao modelo com relação aos dados.

A distribuição escolhida para o ajuste é a distribuição Gama por motivos que já foram descritos ao analisar o gráfico e o histograma da série.

Foi utilizado o algoritmo de Metropolis (Seção 4.3), com um número total de 22000 iterações, 2000 valores iniciais desprezados (*burn-in*) e com salto de tamanho 10 que resultou em uma amostra de tamanho 2000 para cada parâmetro. Assim como no método Escore de Fisher, foram ajustados modelos com ordens $p = 1, \dots, 4$ e $q = 0$. Os resultados para $p = 4$ foram omitidos já que não foi possível obter convergência pelo método.

A Tabela 5.17 traz um resumo das distribuições marginais a posteriori para os modelos ajustados com seus respectivos parâmetros estimados.

Tabela 5.17: Resumo das distribuições marginais *a posteriori* para o modelo GAR Gama(p).

(p, q)		Média	D.P.	<i>I.Cred. de 95%</i>	z	<i>BICM</i>
(1,0)	$\hat{\alpha}_0$	0.7823	0.0972	[0.5892 ; 0.9699]	0.6512	-94.83
	$\hat{\phi}_1$	0.5626	0.0453	[0.4778 ; 0.6562]	-0.0421	
	$\hat{\varphi}$	0.1421	0.0022	[0.1377 ; 0.1471]	-1.0162	
(2,0)	$\hat{\alpha}_0$	0.6108	0.1053	[0.4058 ; 0.8161]	1.0099	-101.75
	$\hat{\phi}_1$	0.3307	0.0571	[0.2197 ; 0.4450]	-0.4072	
	$\hat{\phi}_2$	0.3136	0.0565	[0.2001 ; 0.4193]	-0.0882	
	$\hat{\varphi}$	0.1212	0.0016	[0.1180 ; 0.1247]	-0.3485	
(3,0)	$\hat{\alpha}_0$	0.4296	0.0747	[0.2821 ; 0.5733]	-0.7419	-121.57
	$\hat{\phi}_1$	0.1338	0.0349	[0.0644 ; 0.2072]	-0.6686	
	$\hat{\phi}_2$	0.1353	0.0343	[0.0645 ; 0.2063]	0.5883	
	$\hat{\phi}_3$	0.4813	0.0473	[0.3882 ; 0.5745]	0.1113	
	$\hat{\varphi}$	0.0827	0.0007	[0.0812 ; 0.0842]	-0.2903	

De acordo com os resultados na Tabela 5.17, nota-se que os três modelos produziram resultados condizentes, ou seja, nenhuma estimativa dos parâmetros foi não-significativa. Ainda, de acordo com a Seção 4.4, o modelo que apresentar menor BICM será o modelo selecionado. Portanto, o modelo escolhido com a análise foi o GAR Gama(3).

Além disso, para todos os ajustes realizados as estatísticas z de Geweke ficaram dentro do valor considerado apropriado. Logo, há indícios de que houve convergência.

Os gráficos de ajuste e dispersão para o modelo selecionado, e também um resumo mais detalhado das estimativas obtidas estão a seguir.

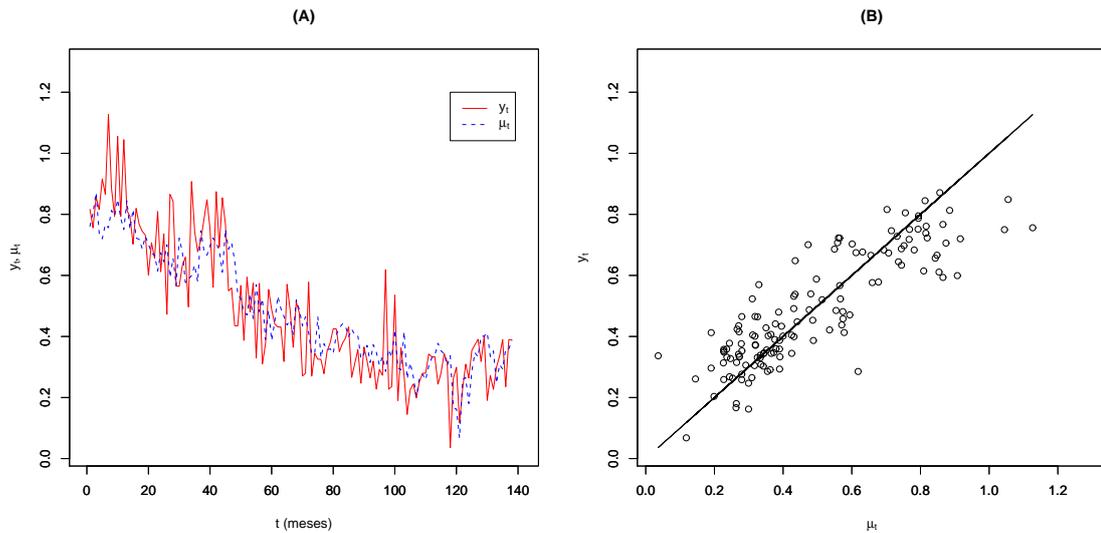


Figura 5.12: (A) Série e média ajustada via MCMC; (B) Dispersão dos dados ajustados.

Pela análise da série ajustada na Figura 5.12, pode-se concluir que há evidências de que o modelo escolhido para realizar o ajuste é o mais apropriado.

Tabela 5.18: Resumo das distribuições marginais *a posteriori* para o modelo GAR Gama(3).

	Média	Mediana	Moda	D.P.	<i>I.Cred. de 95%</i>
$\hat{\alpha}_0$	0.4296	0.4313	0.4315	0.0747	[0.2821 ; 0.5733]
$\hat{\phi}_1$	0.1338	0.1330	0.1231	0.0349	[0.0644 ; 0.2072]
$\hat{\phi}_2$	0.1353	0.1349	0.1291	0.0343	[0.0645 ; 0.2063]
$\hat{\phi}_3$	0.4813	0.4810	0.4988	0.0473	[0.3882 ; 0.5745]
$\hat{\varphi}$	0.0827	0.0827	0.0827	0.0007	[0.0812 ; 0.0842]

Previsão

A previsão para o conjunto de dados aqui apresentado será feita conforme descrito na Seção 4.5, utilizando o modelo GAR Gama(3) que foi o modelo que melhor se ajusta ao conjunto. Como na previsão clássica, o tamanho do passo de previsão será de $h = 5$ e também será utilizado o Erro Médio Percentil Absoluto (MAPE) dado em (3.6), para verificar a qualidade da previsão.

Tabela 5.19: Previsão para $h = 5$ meses

Nº da observação	Valor original	Valor previsto
137	0.3354	0.3535
138	0.3897	0.3534
139	0.2355	0.3942
140	0.3895	0.3974
141	0.3895	0.4041

De acordo com a Tabela 5.19, nota-se que os valores previstos conseguem captar a leve tendência crescente da parte final da série, importante pois a série se mostra decrescente na maior parte do tempo com uma leve mudança apenas na parte final. O MAPE calculado para avaliar a previsão, ficou em torno de 17.58%, evidenciando que a previsão feita apresenta valores coerentes. A figura 5.11 mostra os valores observados e os valores preditos.

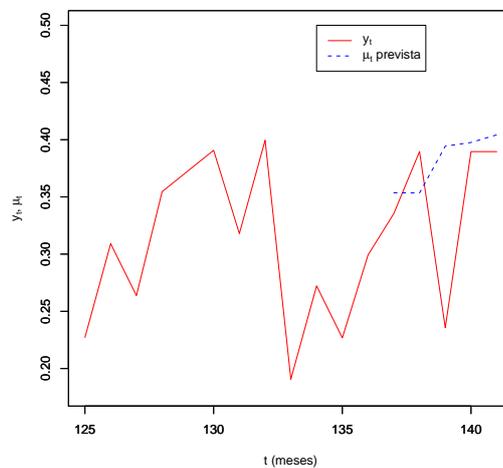


Figura 5.13: Previsão 5 meses a frente.

Pela análise da Figura 5.13, fica claro o que foi explicado no parágrafo anterior e então pode-se concluir que o modelo mostra ter uma capacidade de previsão satisfatória.

5.2.4 Comentários finais sobre as aplicações

Pela análise dos ajustes realizados à série, tanto através dos métodos clássicos como bayesianos, nos forneceram ajustes com uma qualidade bastante boa, inclusive com as estimativas dos parâmetros próximas entre as duas abordagens. Pode-se dizer então que o modelo GARMA consegue, em ambos os ajustes, descrever bem o comportamento da série.

Capítulo 6

Conclusões e considerações finais

Os modelos GARMA ampliam ainda mais os limites impostos sobre alguns modelos clássicos de séries temporais, como é o caso do modelo ARIMA (p, d, q) , por exemplo. Isto traz como benefício, a quantidade maior de séries que podem ser utilizadas, por exemplo, séries para dados de contagem (que possuem distribuições discretas), séries com assimetria (como é o caso da série real tratada na Seção 5.2) e, por vezes, séries financeiras também, como pode ser visto no Apêndice B.

Para os exemplos vistos durante este trabalho, pode-se tirar como conclusão que essa classe de modelos consegue ser bastante eficiente no ajuste da série, possuindo uma estimativa para os parâmetros bem próximas dos seus respectivos valores reais. Isso pode ser reforçado pelas boas propriedades assintóticas apresentadas no estudo de simulação realizado para tais modelos, mostrando que mesmo com amostras pequenas ($n = 100$), os estimadores são assintoticamente eficientes. Isto também é verificado através do vício, uma vez que este tende a zero conforme o tamanho da amostra aumenta.

A abordagem bayesiana, proposta ao longo do texto, também mostra ser uma boa alternativa uma vez que nos estudos realizados com séries simuladas as estimativas dos parâmetros ficaram muito próximas dos valores reais. Uma proposta a ser feita em trabalhos futuros é a mudança das distribuições a priori, a fim de verificar o comportamento e a performance das novas distribuições a posteriori.

Um grande problema encontrado no ajuste dos modelos, foi a não convergência

da parte MA do modelo quando utilizado o método Escore de Fisher. Com o uso de distribuições a priori não informativas também não foi possível resolver esta questão. Uma possibilidade está no uso da priori de Jeffreys que, como mencionado, pode fazer parte de um trabalho futuro para tentar resolver essa questão.

Uma maneira de flexibilizar ainda mais os modelos GARMA está no uso de transformações, já que estas ampliam mais as possibilidades de utilização de tais modelos, pois permitem séries cujas distribuições apresentam caudas pesadas e de distribuições que não pertençam à família exponencial possam ser modeladas após uma transformação adequada.

Em geral, é utilizada a transformação potência de Box-Cox (1964) para séries temporais positivas, dada por

$$Y_t^{(\lambda)} = \begin{cases} \frac{(Y_t^\lambda - 1)}{\lambda}, & \text{se } \lambda \neq 0; \\ \log(Y_t), & \text{se } \lambda = 0, \end{cases}$$

ou então a transformação de Manly (1976), utilizada para séries negativas e também podem ser utilizadas transformações para dados de contagem, dados binários e séries temporais categóricas. Essa transformação é dada por

$$Y_t^{(\lambda)} = \begin{cases} \frac{e^{\lambda Y_t} - 1}{\lambda}, & \lambda \neq 0 \\ \log(Y_t), & \lambda = 0 \end{cases}$$

Um dos maiores interesses no uso de transformações está em, ao utilizar um modelo GARCH em que ε_t sejam não normais, por exemplo a distribuição *t-Student*, não será possível construir um modelo GARMA Gama GARCH (Apêndice B) para esta série, já que tal distribuição não pertence a família exponencial. Não conseguimos também determinar qual distribuição de ν_t e, conseqüentemente, a de $y_t|\mathbf{H}_t$.

Para esse caso, uma possível proposta do trabalho será encontrar uma transformação que, aplicada a série original, a variável de interesse passe a ter uma distribuição pertencente à família de distribuições exponencial para que, com isso, possam ser utilizadas as propriedades do modelo GARMA para o ajuste do modelo.

Referências Bibliográficas

- Akaike, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- Benjamin, M. A., Rigby, R. A. & Stasinopoulos, D. M. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98:214–223, 2003.
- Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.
- Box, G. & Cox, D. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252, 1964.
- Box, G., Jenkins, G. & Reinsel, G. *Time series analysis: forecasting and control*. John Wiley, 4th edition, 2008.
- Brockwell, P. & Davis, R. *Introduction to time series and forecasting*. Springer Verlag, 2002. ISBN 0387953515.
- Chib, S. & Greenberg, E. Understanding the Metropolis-Hastings algorithm. *American Statistician*, 1:327–335, 1995.
- Cordeiro, G., Cribari-Neto, F., Aubin, E. & Ferrari, S. Bartlett corrections for one-parameter exponential family models. *Journal of Statistical Computation and Simulation*, 53(3):211–231, 1995.
- Cordeiro, G. M. & Andrade, M. G. Transformed generalized linear models. *Journal of Statistical Planning and Inference*, 138(9):2970–2987, 2009.
- Ehlers, R. S. *Introdução à inferência Bayesiana*, 2007.

- Engle, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4): 987–1007, 1982.
- Fokianos, K. & Kedem, B. Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis*, 25(2):173–197, 2004.
- Gelman, A. & Rubin, D. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.
- Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, (4):625–631, 1992.
- Li, W. K. Times series model based on generalized linear models: Some further results. *Biometrics*, 50:506–511, 1994.
- Manly, B. Exponential data transformations. *The Statistician*, pages 37–42, 1976.
- McCullagh, P. & Nelder, J. *Generalized Linear Models*. Monographs on Statistics and Applied Probability. London: Chapman and Hall, second edition, 1989.
- Morettin, P. *Econometria Financeira: Um Curso em Séries Temporais Financeiras*. São Paulo: Blücher, 2008.
- Nelder, J. A. & Wedderburn, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society*, 135:370–384, 1972.
- Paulino, C., Turkman, M. & Murteira, B. *Estatística bayesiana*. 2003.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Raftery, A., Newton, M., Satagopan, J. & Krivitsky, P. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Memorial Sloan-Kettering Cancer Center Department of Epidemiology and Biostatistics Working Paper Series*, page 6, 2006.

Schwarz, G. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

Apêndice A

Geração dos dados

A teoria de Modelos Lineares Generalizados diz que, cada observação da variável resposta tem uma distribuição, contando com uma média e uma variância. A média é referente a cada observação, no entanto a variância é comum a todas as observações. Para ilustrar este fato, suponha que temos um conjunto de observações y_t , $t = 1, \dots, n$, e que esses dados são provenientes de uma distribuição normal com média μ_t e variância σ^2 . Então $y_t \sim N(\mu_t, \sigma^2)$, $t = 1, \dots, n$.

A mesma coisa acontece com os modelos GARMA, em que cada observação terá origem de uma média μ_t , porém todas compartilham de uma mesma variância, definida anteriormente como φ . Como visto no Capítulo 2, a média μ_t está associada a um preditor linear, dado por (2.3), e através deste preditor são geradas médias μ_t , para, a partir delas, gerar as observações y_t de acordo com a distribuição desejada.

Suponha por exemplo, o modelo GARMA Normal com $p = 2$ e $q = 0$. Então o preditor linear é dado por $\mu_t = \alpha_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2}$, ou generalizando, $g(\mu_t) = \alpha_0 + \phi_1 g(y_{t-1}) + \phi_2 g(y_{t-2})$, em que $g(\cdot)$ é a função de ligação conhecida. Os valores das médias são gerados atribuindo valores, neste caso, para σ^2 , α_0 , ϕ_1 , ϕ_2 , y_2 e y_1 . Com o valor das médias que foram geradas, uma nova observação é gerada de uma distribuição Normal de média μ_t calculada anteriormente, e variância σ^2 , ou seja, para o exemplo acima, $y_3 \sim N(\mu_3, \sigma^2)$, com $\mu_3 = \alpha_0 + \phi_1 g(y_2) + \phi_2 g(y_1)$.

O algoritmo abaixo exemplifica o processo de simulação de dados que foi descrito

para um modelo GARMA(p,q) mais geral. Todo o processo de simulação e aplicação foi realizado no software livre R Development Core Team (2010).

1. Atribua valores para os parâmetros φ , α_0 , ϕ e θ , e também para observações iniciais y_i e μ_i , $i = 1, \dots, \max(p, q)$;
2. Calcular a média usando $g(\mu_t) = \alpha_0 + \sum_{i=1}^p \phi_i g(y_{t-i}) + \sum_{j=1}^q \theta_j (g(y_{t-j}) - g(\mu_{t-j}))$;
3. Gerar $y_t \sim P(\mu_t, \varphi)$, em que P é uma distribuição pertencente à família exponencial dada em (2.1);
4. Repetir os passos acima até que o número desejado de observações seja atingido.

Com a geração de dados feitas acima, foram feitos os estudos de simulação e também os ajustes clássicos e bayesianos.

Apêndice B

Modelos GARCH para séries financeiras

Na modelagem de séries financeiras, os modelos ARMA não são adequados para descrever o comportamento das mesmas, já que essas apresentam a variância condicional evoluindo no tempo. É necessária uma classe de modelos não-lineares para descrever este comportamento, e uma das mais utilizadas para este fim são os modelos ARCH ("autoregressive conditional heteroscedasticity") introduzido por Engle (1982), e sua generalização, o modelo GARCH, ou Generalized ARCH, introduzido por Bollerslev (1986). Esses modelos são não-lineares no que se refere à variância e descrevem bem a volatilidade de séries financeiras e por isso são largamente utilizados ainda hoje para tal modelagem.

O principal objetivo aqui será modelar a *volatilidade* de uma série financeira, que é a variância condicional da série. Geralmente, a volatilidade que queremos modelar é de uma série de retornos de um ativo financeiro P_t , que é dada por

$$r_t = \log \left(\frac{P_t}{P_{t-1}} \right) \approx \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Sejam

$$\mu_t = E(r_t | \mathcal{F}_{t-1})$$

$$h_t = Var(r_t | \mathcal{F}_{t-1}).$$

a média e a variância condicional de r_t , sendo \mathcal{F}_{t-1} a informação disponível até o instante $t - 1$ que consideramos ser $\{r_{t-1}, r_{t-2}, \dots\}$. Vamos supor também que $\mu_t = 0$, e assim $h_t = E(r_t^2 | \mathcal{F}_{t-1})$.

B.0.5 Modelo GARCH

Sugerida por Bollerslev (1986), o modelo GARCH surge como uma generalização para o modelo ARCH pois, assim como os modelos ARMA, o GARCH pode ser uma ferramenta mais parcimoniosa para descrever a volatilidade com menos parâmetros do que em um modelo ARCH.

Definição B.1 Um modelo $GARCH(p,q)$ é definido por

$$\begin{aligned} r_t &= \sqrt{h_t} \varepsilon_t, \\ h_t &= \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}, \end{aligned} \tag{B.1}$$

em que ε_t é uma variável aleatória independente e identicamente distribuída com média 0 e variância 1, (geralmente distribuição normal ou t de Student), $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\sum_{i=1}^{p^*} \alpha_i + \beta_i < 1$, $p^* = \max(p, q)$.

Considerando agora, que \mathbf{H}_t seja o conjunto de informações passadas, ou seja, $\mathbf{H}_t = \{r_{t-1}, \dots, r_1, h_t, \dots, h_1\}$ e, chamando $v_t = r_t^2 - h_t = h_t(\varepsilon_t^2 - 1)$, tem-se que r_t^2 seguirá um modelo ARMA (p^*, q) , $p^* = \max(p, q)$.

De fato, seja $v_t = r_t^2 - h_t = h_t(\varepsilon_t^2 - 1)$. Então

$$\begin{aligned} r_t^2 &= v_t + h_t \\ &= v_t + \alpha_0 + \sum_{j=1}^p \alpha_j r_{t-j}^2 + \sum_{j=1}^q \beta_j h_{t-j} \\ &= v_t + \alpha_0 + \sum_{j=1}^p \alpha_j r_{t-j}^2 + \sum_{j=1}^q \beta_j (r_{t-j}^2 - v_t) \\ &= v_t + \alpha_0 + \sum_{j=1}^{p^*} (\alpha_j + \beta_j) r_{t-j}^2 - \sum_{j=1}^q \beta_j v_t \end{aligned}$$

em que $p^* = \max(p, q)$. Assim,

$$r_t^2 = \varphi_0 + \sum_{j=1}^{p^*} \varphi_j r_{t-j}^2 + v_t - \sum_{j=1}^q \beta_j v_t$$

segue um modelo ARMA (p^*, q) , com $\varphi_0 = \alpha_0$ e $\varphi_j = \alpha_j + \beta_j$.

É importante lembrar que uma das vantagens de poder escrever o modelo GARCH na forma de um modelo ARMA, é de usar as ferramentas deste último para encontrar, por exemplo, a ordem dos modelos e o cálculo de previsões.

Há muitas variações dos modelos GARCH, como por exemplo o modelo EGARCH (“Exponential GARCH”) e o modelo TARCH (“Threshold ARCH”), este último um caso particular do modelo ARCH, na qual a volatilidade segue uma forma funcional.

Tem-se também os modelos de volatilidade estocástica (MVE) que, diferentemente dos modelos ARCH e GARCH que supõem que a volatilidade depende dos retornos passados, não fazem essa suposição. Este modelo supõe que a volatilidade presente depende de valores passados da mesma, mas que é independente dos retornos passados. Por não se tratar do objetivo deste trabalho, não será aprofundado mais do que o exposto acima sobre estas extensões. Para maiores detalhes ver, por exemplo, Morettin (2008).

B.0.6 O modelo GARMA GARCH

Um caso especial do modelo Gama GARMA é uma forma reparametrizada do modelo GARCH (Bollerslev, 1986).

Se \mathbf{H}_t é o conjunto de informações passadas, ou seja, $\mathbf{H}_t = \{r_{t-1}, \dots, r_1, h_t, \dots, h_1\}$, e fazendo $v_t = r_t^2 - h_t = h_t(\varepsilon_t^2 - 1)$, tem-se que r_t^2 seguirá um modelo ARMA (p^*, q), $p^* = \max(p, q)$.

Suponha agora que $\varepsilon_t \sim N(0, 1)$ e considere a transformação $v_t = h_t(\varepsilon_t^2 - 1)$. Então se $\varepsilon_t \sim N(0, 1)$, $\varepsilon_t^2 \sim \chi_{(1)}^2$. Sabendo que $v_t = h_t(\varepsilon_t^2 - 1)$ e considerando $Y_t = v_t$, tem-se

$$\begin{aligned} f_Y(y_t) &= f_{\chi_{(1)}^2} \left(\frac{h_t - y_t}{h_t} \right) \frac{1}{h_t} \\ &= \frac{2^{-1/2}}{\Gamma(1/2)} \left(\frac{y_t + h_t}{h_t} \right)^{-1/2} \exp \left(-\frac{y_t + h_t}{h_t} \right) \frac{1}{h_t}. \end{aligned} \quad (\text{B.2})$$

Reescrevendo a Equação (B.2) de forma conveniente, obtém-se

$$f_Y(y) = \frac{2h_t^{-1/2}}{\Gamma(1/2)} (y_t + h_t)^{-1/2} \exp(y_t + h_t) \frac{1}{2h_t}. \quad (\text{B.3})$$

De (B.3), tem-se que $Y_t = v_t + h_t \sim \text{Gama} \left(\frac{1}{2}, \frac{1}{2h_t} \right)$.

Além do mais, a esperança condicionada aos dados passados de $v_t + h_t$ é dada por

$$E(v_t + h_t | \mathbf{H}_t) = E(v_t | \mathbf{H}_t) + E(h_t | \mathbf{H}_t) = 0 + h_t = h_t.$$

com h_t dado em (B.1).

Daí, pode-se afirmar que, se $y_t = r_t^2$ e tem distribuição Gama com $E(y_t | \mathbf{H}_t) = E(r_t^2 | \mathbf{H}_t) = \mu_t$, então y_t segue um modelo Gama GARCH.

De fato,

$$\begin{aligned} r_t^2 &= \alpha_0 + \sum_{j=1}^p \alpha_j r_{t-j}^2 - \sum_{j=1}^q \beta_j v_{t-j} + v_t \\ &= h_t + v_t. \end{aligned}$$

Calculando a esperança condicionada a \mathbf{H}_t tem-se

$$E(r_t^2 | \mathbf{H}_t) = \alpha_0 + \sum_{j=1}^p \alpha_j r_{t-j}^2 - \sum_{j=1}^q \beta_j v_{t-j}.$$

Como $E(r_t^2 | \mathbf{H}_t) = \mu_t$, então

$$\begin{aligned} \mu_t &= \alpha_0 + \sum_{j=1}^p \alpha_j r_{t-j}^2 - \sum_{j=1}^q \beta_j v_{t-j} \\ &= \alpha_0 + \sum_{j=1}^p \alpha_j r_{t-j}^2 - \sum_{j=1}^q \beta_j (r_{t-j}^2 - \mu_{t-j}) \\ &= \varphi_0 + \sum_{j=1}^{p^*} \varphi_j r_{t-j}^2 - \sum_{j=1}^q \beta_j \mu_{t-j}, \end{aligned} \tag{B.4}$$

com $p^* = \max(p, q)$.

O preditor linear então é dado por (B.4) cuja função de ligação é a função identidade. Desta forma, se $y_t | \mathbf{H}_t$ tem distribuição Gama e o preditor linear é dado por (B.4), define-se o modelo GARMA Gama GARCH.

Pode-se então, ter no modelo GARMA uma outra alternativa para o tratamento de séries financeiras, uma vez que se as inovações tiverem forma da distribuição normal o uso do modelo GARMA Gama com função de ligação identidade será equivalente ao modelo GARCH.