2
Generalized Autoregressive Moving Average Models (GARMA)

Este capítulo está fundamentado no artigo de Benjamin, Rigby e Stasinopoulos (2003) [1] o qual tem como objetivo tratar a extensão dos modelos ARMA (Autoregressive Moving Average Models) com distribuição condicional gaussiana para distribuições condicionais não-Gaussianas, pertencentes à família exponencial. Os modelos GARMA de Benjamin, Rigby e Stasinopoulos[1] são extensões naturais dos modelos não-Gaussianos desenvolvidos por Zeger e Qaquish (1988) [23] e Li (1994) [13]. Conforme anteriormente colocado, neste trabalho estaremos focados nas especificações GARMA que podem ser utilizadas para modelar séries de dados de contagem que apresentam dependência temporal, a saber, os modelos GARMA com distribuições condicionais Poisson e Binomial Negativa.

Os modelos GARMA são definidos neste trabalho com a mesma notação e arcabouço utilizados nos GLM (*Generalized Linear Models*) propostos por McCullagh e Nelder (1989) [14] para observações independentes. A diferença fundamental entre as duas abordagens é que enquanto em McCullagh e Nelder a distribuição definida é a marginal, nos modelos de Benjamin, Rigby e Stasinopoulos a parte aleatória do modelo é especificada pela distribuição condicional tornando assim possível o tratamento de dados (amostras) dependentes. Assim sendo, é conveniente introduzir uma pequena descrição dos GLM, antes de abordarmos os modelos GARMA.

2.1
Generalized Linear Models (GLM)

Trata-se da generalização do modelo de regressão linear e gaussiano de forma a adequá-lo para a modelagem de variáveis de resposta independentes que apresentem características explícitas de não-normalidade, tais como variáveis continuas com assimetria e dados de contagem.

Consideremos primeiro o modelo de regressão linear clássico onde:

$y_{(nx1)}$ é o vetor de observações;

$\mu_{(nx1)}$ é a parte sistemática ou explicativa do modelo;

x_j são as variáveis explicativas (regressores não estocásticos), onde

j =1,2, ... ,p;

No arcabouço dos GLM, o modelo de regressão é especificado através de três componentes, a saber:

- Parte aleatória:

$$p(y) \text{ é } NID(\mu, \sigma^2 I); \quad \textbf{(1)}$$

- Parte explicativa:

$$\eta = \sum_{j=1}^{p} \beta_j x_j \quad \textbf{(2)}$$

onde η é denominado o preditor linear;

- Função de ligação: componente que faz a ligação entre a parte aleatória e a parte explicativa do modelo de regressão:

$$\mu = \eta \quad \textbf{(3)}$$

Para o caso generalizado temos as seguintes diferenças em relação ao caso linear clássico:

- componente aleatória: a distribuição vem de uma família exponencial, a qual engloba como casos particulares, além da Gaussiana: Binomial, Poisson, Binomial Negativa, Gama e Inversa Gaussiana.

- componente de ligação: a função de ligação não necessita ser a identidade. A escolha da função de ligação adequada vai depender da distribuição escolhida podendo ser qualquer função monotônica duplamente diferenciável:

$$g(\mu) = \eta \quad \text{onde} \quad g(\cdot) \text{ é a função de ligação.}$$

A estimação dos modelos é efetuada por máxima verossimilhança, a qual é demonstrado ser equivalente a um esquema de mínimos quadrados ponderados reiterados. Dada a breve descrição da modelagem GLM, agora será apresentada a definição do modelo GARMA.

2.2
Definição do modelo GARMA

Nos modelos GARMA a distribuição condicional de cada observação y_t para cada $t = 1, \ldots, n$ dado o conjunto de informações passadas $\underline{H}_t$ onde:

$$\underline{H}_t = \{\underline{x}_t, \ldots, \underline{x}_1, y_{t-1}, \ldots, y_1, \mu_{t-1}, \ldots, \mu_1\}; \qquad \textbf{(4)}$$

é considerada pertencente a uma mesma família exponencial abaixo representada:

$$f(y_t / \underline{H}_t) = \exp\left\{ \frac{y_t \vartheta_t - b(\vartheta_t)}{\varphi} + d(y_t, \varphi) \right\}; \qquad \textbf{(5)}$$

onde ϑ_t é o parâmetro canônico da família exponencial, φ é o parâmetro de escala, $b(\cdot)$ e $d(\cdot)$ são funções específicas que definem uma família exponencial específica e $\underline{x}$ é o vetor de variáveis explicativas de tamanho k.

Como no caso GLM, a média condicional do processo μ_t se relaciona com o preditor η_t pela função de ligação $g(\cdot)$, ou seja, $g(\mu_t) = \eta_t$ sendo $g(\cdot)$ duplamente diferenciável e monotônica.

Diferentemente do caso padrão GLM, aqui, o preditor η_t possui uma componente adicional, τ_t que provê a parte relativa aos termos autoregressivos e médias móveis, bem como também inclui no preditor valores passados das variáveis explicativas presentes no modelo. Portanto, teremos:

$$g(\mu_t) = \eta_t = \underline{x}_t' \underline{\beta} + \tau_t \; ; \qquad \textbf{(6)}$$

com

$$\tau_t = \sum_{j=1}^{p} \phi_j A\left(y_{t-j}, x_{t-j}, \underline{\beta}\right) + \sum_{j=1}^{q} \theta_j M\left(y_{t-j}, \mu_{t-j}\right) \qquad \textbf{(7)}$$

onde A e M são funções que representam os termos autoregressivos e médias móveis respectivamente, com parâmetros associados $\underline{\phi}' = \left(\phi_1, \ldots, \phi_p\right)$ e $\underline{\theta}' = \left(\theta_1, \ldots, \theta_q\right)$.

O modelo apresentado acima é muito geral e será particularizado da mesma forma presente no artigo de Benjamin, Rigby e Stasinopoulos (2003) [1]. Abaixo está descrita a nova expressão para o preditor η_t feitas as devidas alterações no termo τ_t:

$$\eta_t = \underline{x}_t' \underline{\beta} + \sum_{j=1}^{p} \phi_j \left\{ g\left(y_{t-j}\right) - \underline{x}_{t-j}' \underline{\beta} \right\} + \sum_{j=1}^{q} \theta_j \left\{ g\left(y_{t-1}\right) - \eta_{t-1} \right\} \qquad \textbf{(8)}$$

.

Logo, juntando essa equação à equação de $f\left(y_t / \underline{H}_t\right)$ podemos definir o modelo GARMA(p,q):

$$f\left(y_t / \underline{H}_t\right) = \exp\left\{ \frac{y_t \vartheta_t - b(\vartheta_t)}{\varphi} + d\left(y_t, \varphi\right) \right\} \quad \textbf{(9)}$$

$$g(\mu_t) = \eta_t = \underline{x}_t^{'} \underline{\beta} + \sum_{j=1}^{p} \phi_j \left\{ g\left(y_{t-j}\right) - \underline{x}_{t-j}^{'} \underline{\beta} \right\} + \sum_{j=1}^{q} \theta_j \left\{ g\left(y_{t-1}\right) - \eta_{t-1} \right\} \quad \textbf{(10)}$$

.

Da parametrização do modelo GARMA acima descrito e usando os seguintes resultados já conhecidos e abaixo enunciados, é possível chegar nas expressões da média condicional de y_t dado $\underline{H}_t$, $E\left[y_t / \underline{H}_t\right]$ e da variância condicional de y_t dado $\underline{H}_t$, $Var\left(y_t / \underline{H}_t\right)$. Detalhes deste desenvolvimento podem ser encontrados no apêndice [I].

Utilizando as relações:

$$E\left[\frac{\partial \log f\left(y_t / \underline{H}_t\right)}{\partial \vartheta_t} \right] = 0 \quad \textbf{(11)}$$

$$E\left[\frac{\partial^2 \log f\left(y_t / \underline{H}_t\right)}{\partial \vartheta_t^2} \right] + E\left[\left(\frac{\partial \log f\left(y_t / \underline{H}_t\right)}{\partial \vartheta_t} \right)^2 \right] = 0 \quad \textbf{(12)}$$

chega-se à:

- média condicional $E\left[y_t\right] = \dot{b}\left(\vartheta_t\right)$ **(13)**

- variância condicional $Var\left(y_t\right) = \varphi \ddot{b}\left(\vartheta_t\right)$ **(14)**

Nesta dissertação iremos explorar duas especificações dos modelos GARMA apresentadas no artigo de Benjamin, Rigby e Stasinopoulos, a saber, o modelo GARMA-Poisson e o modelo GARMA-Binomial Negativa (GARMA-NB), os quais serão apresentados a seguir.

2.3
Modelo GARMA-POISSON

O modelo GARMA-Poisson consiste no modelo GARMA onde a função de distribuição de probabilidade utilizada para y_t dado $\underline{H}_t$ é a distribuição Poisson. Para este modelo a função de ligação escolhida foi a logarítmica pois está provê valores não negativos para $\mu_t = g^{-1}(\eta_t)$ não importando os valores atribuídos a η_t. Então, fazendo uso das considerações enunciadas podemos definir o modelo GARMA-Poisson:

$$f(y_t / \underline{H}_t) = \frac{\mu_t^{y_t} e^{-\mu_t}}{y_t!} \quad \textbf{(15)}$$

$$\log(\mu_t) = \underline{x}_t' \underline{\beta} + \sum_{j=1}^{p} \phi_j \left\{ \log(y_{t-j}^*) - \underline{x}_{t-j}' \underline{\beta} \right\} + \sum_{j=1}^{q} \theta_j \left\{ \log\left(\frac{y_{t-j}^*}{\mu_{t-1}} \right) \right\} \quad \textbf{(16)}$$

onde é conhecido que : $E[y_t / \underline{H}_t] = Var[y_t / \underline{H}_t] = \mu_t$ **(17)**

onde $y_{t-j}^* = \max(y_{t-j}, c)$ para $0 < c < 1$. O artifício de usar y_{t-1}^* ao invés de y_{t-j} se mostrou necessário para substituir os valores em que $y_{t-j} = 0$ por c a fim de permitir o uso da função logarítmica como função de ligação para modelagem de séries que possuam algumas observações iguais a zero.

São casos particulares da modelagem GARMA-Poisson os modelos puramente autoregressivos ($\theta_j = 0$ para $j = 1,2,\ldots,q$) estudados por Zeger e Qaqish (1988) [23] e também os modelos puramente médias móveis, ($\phi_j = 0$ para $j = 1,2,\ldots,p$) estudados por Li (1994) [13].

Algumas outras formas de modelagem foram estudadas por Davis et al. (1999) [6] e Brumback et al. (2000) [5]. Davis discute em seus estudos mudanças na expressão do preditor utilizando os resíduos de Pearson para os termos médias móveis, enquanto Brumback definiu um modelo similar ao GARMA-Poisson chamado TGLM (*Transitional Generalized Linear Models*) propondo quatro diferentes formas de se definir o preditor.

2.4
Modelo GARMA-BINOMIAL NEGATIVA (GARMA-NB)

No modelo GARMA-Binomial Negativa temos que a distribuição de probabilidade condicional utilizada é a distribuição Binomial Negativa. A parametrização utilizada e abaixo enunciada é a mesma presente em Benjamin, Rigby e Stasinopoulos (2003) [1].

$$f\left(y_t / \underline{H}_t\right) = \frac{\Gamma\left(y_t + k\right)}{\Gamma(k)\Gamma\left(y_t + 1\right)} \left\{\frac{\mu_t}{\mu_t + k}\right\}^{y_t} \left\{\frac{k}{\mu_t + k}\right\}^{k} \quad \textbf{(18)}$$

para $y_t = 0,1,2,\ldots$ e $k > 0$ onde k é o parâmetro de dispersão. Nesta parametrização a média condicional e a variância condicional dado y_t podem ser encontrados pelas respectivas expressões:

$$E\left[y_t / \underline{H}_t\right] = \mu_t \quad \textbf{(19)}$$

$$Var\left[y_t / \underline{H}_t\right] = \mu_t + \frac{\mu_t^2}{k} \quad \textbf{(20)}$$

onde a expressão para log(μ_t) é a mesma utilizada no modelo GARMA-Poisson:

$$\log\left(\mu_t\right) = \underline{x}_t^{'}\underline{\beta} + \sum_{j=1}^{p}\phi_j\left\{\log\left(y_{t-j}^{*}\right) - \underline{x}_{t-j}^{'}\underline{\beta}\right\} + \sum_{j=1}^{q}\theta_j\left\{\log\left(\frac{y_{t-j}^{*}}{\mu_{t-1}}\right)\right\} \quad \textbf{(21)}$$

Verifica-se facilmente para a distribuição Binomial Negativa que:

$$Var[y_t / \underline{H}_t] > E[y_t / \underline{H}_t] \qquad \textbf{(22)}$$

ou seja, o modelo GARMA-NB apresenta superdispersão em relação ao modelo GARMA-Poisson. Detalhes dos cálculos da média e da variância condicionais bem como a expressão de $f(y_t / \underline{H}_t)$ para o caso GARMA-Negativa Binomial na forma padrão da família exponencial em questão encontram-se no apêndice (II).

2.5
Estimação do vetor de hiperparâmetros

Para estimar o vetor de parâmetros desconhecidos $\underline{\gamma} = (\underline{\beta}', \underline{\phi}', \underline{\theta}')$ de um modelo GARMA utiliza-se o logaritmo da função de verossimilhança, a qual deve ser maximizada:

$$l\left(\underline{\gamma}; y_t, y_{t-1}, \ldots, y_1\right) = \ln\left[L\left(\underline{\gamma}; y_t, y_{t-1}, \ldots, y_t\right)\right] \qquad \textbf{(23)}$$

Tal como nos GLM esta maximização irá exigir a utilização de métodos numéricos. Para o modelo GARMA-Poisson esta função tem a seguinte expressão:

$$l\left(\underline{\gamma}; y_\tau, y_{\tau+1}, \ldots, y_T\right) = \log\left[\prod_{t=\tau}^{T} \frac{\mu_t^{y_t} e^{-\mu_t}}{y_t!}\right] \qquad \textbf{(24)}$$

$$l\left(\underline{\gamma}; y_\tau, y_{\tau+1}, \ldots, y_T\right) = \sum_{t=\tau}^{T} \left[y_t \mu_t - \mu_t - \log(y_t!)\right] \qquad \textbf{(25)}$$

Para o modelo GARMA-NB a expressão é dada por:

$$l\left(\underline{\gamma}; y_\tau, y_{\tau+1}, \ldots, y_T\right) = \log\left[\prod_{t=\tau}^{T} \frac{\Gamma(y_t + k)}{\Gamma(k)\Gamma(y_t + 1)} \left\{\frac{\mu_t}{\mu_t + k}\right\}^{y_t} \left\{\frac{k}{\mu_t + k}\right\}^{k}\right] \qquad \textbf{(26)}$$

$$l\left(\underline{\gamma}; y_\tau, y_{\tau+1}, \ldots, y_T\right) = \sum_{t=\tau}^{T} \Big[\log\left(\Gamma\left(y_t + k\right)\right) - \log\left(\Gamma(k)\right) - \log\left(\Gamma\left(y_t + 1\right)\right) + \\ + y_t \log\left(\frac{\mu_t}{\mu_t + k}\right) + k\left(\frac{k}{\mu_t + k}\right)\Big] \tag{27}$$

onde τ é o índice da primeira observação não nula (a partir da qual temos distribuições próprias). Para estimar os hiperparâmetros dos modelos estudados foi utilizada a ferramenta *fmincon* presente no *software* Matlab. *Fmincon* é uma ferramenta própria para achar o mínimo de uma função de múltiplas variáveis num subespaço paramétrico bem definido. Logo para estimar os valores dos hiperparâmetros bastou minimizar o negativo da função de log-verossimilhança.

2.6 Estacionariedade dos modelos GARMA

Tal qual nos modelos ARMA Gaussianos, é necessário também investigar as condições de estacionariedade dos modelos GARMA, de forma que inferências adequadas sejam obtidas a partir destes modelos. Uma análise das condições necessárias para garantir estacionariedade dos processos GARMA será abaixo apresentada. Esta análise encontra-se dividida em dois casos. No primeiro caso verifica-se as condições de estacionariedade dos modelos GARMA cuja função de ligação é a identidade. No segundo caso estuda-se a condição de estacionariedade dos modelos GARMA cujas funções de ligação não são identidade. Este último caso é o que nos interessa, pois nossa função de identidade é a logarítmica.

2.6.1 Modelos com função de ligação identidade

São definidas nesta seção condições de estacionariedade para a média e a variância incondicionais bem como a expressão da média e da variância. Para um modelo definido pelas equações (9) e (10) a média incondicional pode ser obtida seguindo os seguintes passos a partir da expressão (10) aplicando-se a função de ligação identidade.

$$\mu_t = \underline{x}_t^{'}\underline{\beta} + \sum_{j=1}^{p}\phi_j\left\{y_{t-j} - \underline{x}_{t-j}^{'}\underline{\beta}\right\} + \sum_{j=1}^{q}\theta_j\left\{y_{t-j} - \eta_{t-1}\right\}. \qquad \textbf{(28)}$$

Definindo

$$y_t = \mu_t + \nu_t \qquad \textbf{(29)}$$

temos que ν_t é o erro de previsão com média e autocovariância iguais a zero. Substituindo a expressão (28**)** em (29**)**,

$$y_t = \underline{x}_t^{'}\underline{\beta} + \sum_{j=1}^{p}\phi_j\left\{y_{t-j} - \underline{x}_{t-j}^{'}\underline{\beta}\right\} + \sum_{j=1}^{q}\theta_j\left\{y_{t-j} - \eta_{t-1}\right\} + \nu_t \qquad \textbf{(30)}$$

e reorganizando a expressão acima, chegamos à equação,

$$y_t - \underline{x}_t^{'}\underline{\beta} = \sum_{j=1}^{p}\phi_j\left\{y_{t-j} - \underline{x}_{t-j}^{'}\underline{\beta}\right\} + \sum_{j=1}^{q}\theta_j\left\{y_{t-j} - \eta_{t-1}\right\} + \nu_t\,. \qquad \textbf{(31)}$$

Fazendo

$$\upsilon_{t-1} = y_{t-1} - \eta_{t-1} \quad \text{e} \quad w_t = y_t - \underline{x}_t^{'}\underline{\beta} \qquad \textbf{(32)}$$

pode-se escrever uma nova equação em função de w_t,

$$w_t = \sum_{j=1}^{p}\phi_j w_{t-1} + \sum_{j=1}^{q}\theta_j \upsilon_{t-1} + \nu_t \qquad \textbf{(33)}$$

ou

$$w_t = \Psi(B)\nu_t \qquad \textbf{(34)}$$

onde:

$$\Psi(B) = \Phi(B)^{-1}\Theta(B) \quad \textbf{(35)}$$

$$\Psi(B) = 1 + \Psi_1 B + \Psi_2 B^2 + \ldots \quad \textbf{(36)}$$

$$\Psi^{(2)}(B) = 1 + \Psi_1^2 B + \Psi_2^2 B^2 + \ldots \quad \textbf{(37)}$$

$$\Phi(B) = 1 - \Phi_1 B - \ldots - \Phi_p B^p \quad \textbf{(38)}$$

$$\Theta(B) = 1 + \Theta_1 B + \ldots + \Theta_q B^q \quad \textbf{(39)}$$

onde $\Phi(B)$ é inversível.

Feitas essas considerações poderemos então calcular o valor esperado incondicional de y_t ao tomarmos o valor esperado dos termos da equação (32**)**.

$$E[y_t] = E[\underline{x}_t^{'}\underline{\beta}] - E[w_t] \quad \textbf{(40)}$$

$$E[y_t] = \underline{x}_t^{'}\underline{\beta} \quad \textbf{(41)}$$

Para o cálculo da variância incondicional, faz-se novamente uso da equação (32) chegando-se então na seguinte expressão:

$$Var[y_t] = \varphi\ E\left[\Psi^2(B)\upsilon(\mu_t)\right] \quad \textbf{(42)}$$

Escrevendo a equação da variância na forma acima facilita o reconhecimento das condições de estacionariedade para a mesma. Visto que $\upsilon(\mu_t)$ depende da distribuição condicional especificada, a expressão da variância incondicional apresentará forma distinta, dependendo do modelo considerado, Poisson ou Binomial Negativa. Para o modelo GARMA-Poisson a variância incondicional será dada por:

$$Var[y_t] = \Psi^{(2)}(1)\beta_0 \quad \textbf{(43)}$$

onde $\Psi^{(2)} = 1 + \sum_{j=1}^{\infty} \psi_j^2$.

Enquanto que para o modelo GARMA-NB a variância incondicional será dada por:

$$Var[y_t] = \Psi^{(2)}(1)\left[\beta_0 + \frac{E[\mu_t^2]}{k}\right], \quad \textbf{(44)}$$

onde k é o parâmetro de dispersão presente na distribuição Binomial Negativa e

$$E[\mu_t^2] = \left[1 + \frac{1}{k} - \frac{\Psi^{(2)}(1)}{k}\right]^{-1} \left\{\beta_0^2 + \left[\Psi^{(2)}(1) - 1\right]\beta_0\right\} \quad \textbf{(45)}$$

sendo $\left[1 + \frac{1}{k} - \frac{\Psi^{(2)}(1)}{k}\right]$ inversível.

2.6.2
Modelos com função de ligação diferente da identidade

Diferentemente dos modelos com função de ligação identidade, os modelos que possuem quaisquer outras funções de ligação diferente da identidade não permitem uma forma direta e tratável para o cálculo das expressões dos momentos incondicionais e nem também uma forma simples de se determinar às restrições paramétricas necessárias para garantir estacionariedade.

Podemos verificar este fato analisando um dos casos mais simples em estudo, os modelos GARMA-Poisson e GARMA-Binomial Negativa, para os quais foram assumidos a função de ligação logaritmo, ou seja $g(\cdot) = \log(\cdot)$.

Como exemplo temos o modelo GARMA apenas com um termo autoregressivo (AR(1)), sem variáveis explicativas e intercepto constante igual a $\underline{x}_t'\beta = \beta_0$. Definindo o modelo:

$$\begin{cases} p(y_t / Y_{t-1}) \sim Poisson(\mu_t) \\ \log(\mu_t) = \beta_0 + \phi_1 y_{t-1} \end{cases} \quad \textbf{(46)}$$

O primeiro momento condicional será dado pela expressão:

$$E\left[y_t / Y_{t-1}\right] = \mu_t = e^{(\beta_0 + \phi_1 y_{t-1})} \quad \textbf{(47)}$$

Tentando calcular o primeiro momento incondicional notamos pela equação (50) que este é intratável.

$$E[y_t] = E[E[y_t / Y_{t-1}]] \quad \textbf{(48)}$$

$$E[y_t] = E\left[e^{(\beta_0 + \phi_1 y_{t-1})}\right] \quad \textbf{(49)}$$

$$E[y_t] = e^{\beta_0} E\left[e^{(\phi_1 y_{t-1})}\right] \quad \textbf{(50)}$$

Apenas para fins comparativos, abaixo estão os mesmos quatro passos anteriores para o mesmo modelo, mas agora com a função de ligação identidade:

$$E[y_t / Y_{t-1}] = \mu_t = \beta_0 + \phi_1 y_{t-1} \quad \textbf{(51)}$$

$$E[y_t] = E[E[y_t / Y_{t-1}]] \quad \textbf{(52)}$$

$$E[y_t] = E[\beta_0 + \phi_1 y_{t-1}] \tag{53}$$

$$E[y_t] = \beta_0 + \phi_1 E[y_{t-1}] \tag{54}$$

Como num processo estacionário $E[y_t] = E[y_{t-1}]$ então chegaremos a uma equação para o primeiro momento condicional para este modelo em questão cuja função de ligação é identidade.

$$E[y_t] = \frac{\beta_0}{1-\phi_1} \tag{55}$$

Conforme pode-se observar, para modelos em que a função de ligação difere da identidade, o cálculo dos momentos incondicionais se torna impraticável impossibilitando a análise direta do espaço paramétrico no qual a estacionariedade é garantida.

No capítulo 3, seguindo o procedimento adotado por Benjamin, Rigby e Stasinopoulos (2003) [1], iremos investigar a região de estacionariedade dos modelos GARMA, pela obtenção, via simulação Monte Carlo, da distribuição incondicional implicada pelos modelos GARMA