

**PROGRAMA INTERINSTITUCIONAL DE
PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar-USP**

DEs-UFSCar e SME-ICMC-USP

**ABORDAGEM BAYESIANA DOS MODELOS
GARMA(p, q) USADOS PARA SÉRIES TEMPORAIS
DE DADOS DE CONTAGEM**

**Marinho G. Andrade
Ricardo S. Ehlers
Adriana S. Philippsen
Marcos H. Cascone**

RELATÓRIO TÉCNICO

TEORIA E MÉTODO – SÉRIE A

**Março/2014
nº 259**

Abordagem Bayesiana dos Modelos GARMA(p,q) usados para Séries Temporais de Dados de Contagem

Marinho G. Andrade, Ricardo S. Ehlers, Adriana S. Philippsen
Marcos H. Cascone
SME- ICMC/USP - São Carlos -
13.560-970 - São Carlos / SP.

16 de março de 2014

Resumo

Este trabalho apresenta uma abordagem Bayesiana para os modelos GARMA usados para as séries temporais de dados de contagem. Foram consideradas as distribuições de Poisson, Binomial e Binomial Negativa. A abordagem proposta apresenta também o cálculo de critérios de seleção de modelos baseados nos critérios, *Bayesian Information Criterion* (BIC), *Deviance Information Criterion* (DIC) e na densidade condicional preditiva ordenada (CPO). Além disto a medida de divergência de Kullback-Leibler é proposta para avaliar a influencia dos pontos extremos nas séries temporais de dados de contagem. São apresentados um estudo com dados simulados e uma análise de uma série real de incidência de Dengue no estado da Paraíba no período de janeiro de 1998 a dezembro de 2005.

1 Introdução

Séries temporais para dados de contagem aparecem em várias áreas de aplicação, tais como o estudos de incidência de certas doenças (Zeger, 1988; Davis *et al.*, 2000); séries da movimentação de preços no mercado financeiro (Liesenfeld *et al.*, 2006); em controle de qualidade (Weiß, 2007), na área de seguros, considerando séries de contagem de reivindicação (Freeland & McCabe, 2004). Para que um modelo estatístico de séries de contagem seja bem sucedido, o modelo deve considerar: (1) uma estrutura de dependência

as vezes bastante acentuada, (2) a superdispersão em relação a média da série e (3) permitir a inclusão de covariáveis.

Os modelos estatísticos para dados de contagem dependentes são classificados em Cox (1981) como modelos dirigidos por parâmetros (*parameter-driven*) e modelos dirigidos por observações (*observation-driven*). Os modelos *parameter-driven* estendem os modelos lineares generalizados, introduzindo na média condicional do processo de contagem, um processo autorregressivo latente. Este processo latente é responsável por introduzir autocorrelação e superdispersão nos modelos. Um importante membro desta classe de modelos é o modelo SAM (*Stochastic Autoregressive Mean*) proposto por Zeger (1988) e posteriormente analisado em Davis *et al.* (1999, 2000). O principal problema dos modelos SAM é que a função de verossimilhança depende de integração de alta dimensão, de forma que a estimação eficiente não é simples e geralmente requer métodos baseados em integração de Monte-Carlo. Um dos métodos considerados para a estimativa de modelos SAM é o método de amostragem por importância eficiente (*Efficient Importance Sampling* (EIS)), como aplicado em Jung *et al.* (2006). Em geral, correlações positivas e negativas dos dados de contagem observados podem ser modelados no contexto dos modelos SAM. Recentes contribuições para os modelos *observation-driven* são apresentadas em Davis & Wu (2009), que considera a distribuição binomial negativa, visando tornar mais flexível a modelagem da superdispersão dos processo de contagem. Em Jung & Tremayne (2011) encontra-se uma extensiva discussão sobre a inferência dos modelos SAM.

Os modelos *observation-driven* usam observações passadas na função da média condicional do processo de contagem observado. Entre os modelos *observation-driven*, os modelos que têm recebido mais atenção são: (1) os modelos inteiros-autorregressivos e de médias móveis (INARMA); (2) os modelos autorregressivos média condicional ou modelos autorregressivos condicional de Poisson, $ACP(p,q)$, que também tem sido referido como o modelo INGARCH(p,q); (3) Os modelo lineares generalizados ARMA, referidos como modelos GLARMA(p,q).

Os modelos INARMA foram introduzidos por Al-Osh & Alzaid (1987) e McKenzie (1988) e são modelos cuja estrutura segue a mesma estrutura dos modelos lineares gaussianos para séries temporais ARMA (ver Du & Li (1991)). Porém, a natureza discreta dos dados de contagem é preservada com a aplicação de um operador estocástico que estabelece a relação entre as contagens passadas $\{y_{t-1}, y_{t-2}, \dots\}$, adicionadas a um processo de inovação independente aceitável, e a observação corrente y_t . Esta classe de modelos tem recebido grande atenção nos últimos anos (ver Weiß (2008) e Jung & Tremayne (2011) e e as referências neles citadas).

A classe do modelos *observation-driven* referida como modelos autorre-

gressivos média condicional foi propostos por Heinen (2003) e a especificação inicial destes modelos são os modelos autorregressivos condicional de Poisson, $ACP(p,q)$. Estes modelos consideram uma estrutura para a média condicional do processo de contagem análoga aos modelos GARCH propostos por Bollerslev (1986). Para assegurar a não-negatividade da média condicional e estacionariedade dos modelos $ACP(p,q)$ são impostas aos parâmetros destes modelos as mesmas restrições impostas aos parâmetros dos modelos $GARCH(p,q)$. Devido à sua analogia com o modelo GARCH padrão, o modelo $ACP(p,q)$ também tem sido referido por Ferland *et al.* (2006) como modelo $INGARCH(p, q)$, onde ‘IN’ abrevia ‘valor inteiro’. Os modelos $INGARCH$ são particularmente bem adequados para a modelagem de processos estacionários de contagens superdispersos. Os modelos $INGARCH$, também apresentam vantagens quando comparados aos modelos $INAR$ com distribuição marginal binomial negativa ou Poisson generalizada devido a estes exigirem um parâmetro adicional, e para modelos de ordem superior a 1, a distribuição conjunta dos operadores estocásticos dos modelos $INAR$ levam a modelos com uma estrutura bastante complexa. Recentes contribuições na classe de modelos $INGARCH$ são dados em Weiß (2009), Fokianos *et al.* (2009) que considera a ergodicidade geométrica e inferência baseada em verossimilhança para o processo autorregressivo, linear e não-linear, de Poisson e Zhu (2011) que considera o modelo binomial negativo $INGARCH$.

Outra classe dos modelos *observation-driven* referida como modelos *Generalized Linear Autoregressive Moving Average*, $GLARMA(p,q)$, foi propostos inicialmente por SHEPHARD (1995) e em seguida apresentado em Davis *et al.* (1999, 2003, 2005). O modelo $GLARMA$ estende a estrutura dos modelo linear generalizado, adicionando ao preditor linear, definido pelo logaritmo natural da média condicional do processo de contagem, um resíduo que tem uma estrutura de médias móveis, $MA(q)$. Este ruído, que segue um modelo $MA(q)$, é uma sequência martingale gerada pela diferença entre os dados observados e a média condicional. Por isto é dada a classificação de *observation-driven* a estes modelos. A extensão para considerar uma estrutura $ARMA(p,q)$ pode ser feita considerando um processo latente $ARMA(p,q)$ tendo como ruídos a sequência de diferenças martingale. Nestes casos, estes modelos são referidos como $GLARMA(p,q)$ (Davis *et al.*, 2003). Esta variável latente introduz no modelo linear generalizado de Poisson, a correlação serial entre as observações, bem como a superdispersão, geralmente presente nas séries de dados de contagem.

Neste trabalho vamos considerar outra classe dos modelos *observation-driven* proposta em Benjamin *et al.* (2003) e referida como modelos $GARMA$. Estes modelos, semelhante aos modelos $GLARMA$, são uma extensão dos Modelos Lineares Generalizados (MLG) 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 não-lineares autorregressivos (AR) e de médias móveis (MA). Neste contexto os modelos GARMA é uma extensão dos modelos ARMA gaussianos, para que permite modelar série temporais com distribuição na família exponencial e que tem o modelo ARMA gaussiano como caso particular. Além disto, é possível tratar os modelos GARMA de forma muito semelhante ao tratamento dado aos MLGs.

Vários autores apresentaram casos particulares do modelo GARMA, por exemplo, em Li (1994) e Fokianos & Kedem (2004) os modelos GARMA com distribuição de Poisson, binomial negativa e gama são apresentados e aplicados a conjunto de dados de série temporal real. Por outro lado, outra classe de modelos para séries temporais de dados de contagem (modelos GLARMA) foram propostos por Davis *et al.* (2000).

Uma grande vantagem dos modelos GARMA é permitir a modelagem de séries para dados de contagens, considerando as distribuições de Poisson, Binomial e Binomial Negativa. Nestes casos a abordagem clássica, em geral, apresenta dificuldades na inferência dos termos de médias móveis. Neste artigo, estamos propondo uma abordagem Bayesiana para inferência dos parâmetros dos modelos GARMA para séries de dados de contagem. A abordagem proposta apresenta também o cálculo de critérios de seleção de modelos baseados nos critérios, *Bayesian Information Criterion* (BIC), *Deviance Information Criterion* (DIC) e na densidade condicional preditiva ordenada (CPO). Além disto a medida de divergência de Kullback-Leibler é proposta para avaliar a influencia dos pontos extremos nas séries temporais de dados de contagem.

São apresentados na seção de aplicação, um estudo com dados simulados e uma análise de uma série real de incidência de Dengue (número de notificações/ 100 habitantes) no estado da Paraíba no período de janeiro de 1998 a dezembro de 2005.

2 Modelo GARMA

Seja $\{y_t, t = 1, \dots, n\}$ uma série temporal. Em um modelo GARMA a distribuição condicional de cada observação y_t , condicionada o conjunto de informações passadas \mathbf{H}_t , denotado por:

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

pertence a família exponencial e pode ser escrita de forma geral como:

$$f(y_t|\mathbf{H}_t) = \exp \left\{ \frac{y_t \theta_t - b(\theta_t)}{\varphi} + c(y_t, \varphi) \right\} \quad (1)$$

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}_t é um vetor de variáveis explicativas externas a série temporal.

Assim como nos modelos lineares generalizados, a média condicional do processo μ_t está relacionada com o preditor η_t por uma função de ligação $g(\cdot)$, isto é, $g(\mu_t) = \eta_t$, sendo $g(\cdot)$ uma função monótona e diferenciável, incluindo um componente adicional, τ_t , que inclui os termos auto regressivos e de médias móveis no preditor linear além dos valores das variáveis explicativas presentes no modelo. Assim, podemos escrever

$$g(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \tau_t \quad (2)$$

em que

$$\tau_t = \sum_{j=1}^p \phi_j \mathcal{A}(y_{t-j}, \mathbf{x}_{t-j}, \boldsymbol{\beta}) + \sum_{j=1}^q \alpha_j \mathcal{M}(y_{t-j}, \mu_{t-j}) \quad (3)$$

com \mathcal{A} e \mathcal{M} funções que representam os termos auto regressivos e de médias móveis, respectivamente, com $\boldsymbol{\phi}' = (\phi_1, \dots, \phi_p)$ e $\boldsymbol{\alpha}' = (\alpha_1, \dots, \alpha_q)$ os vetores de parâmetros auto regressivos e de médias móveis, respectivamente.

Na prática, o preditor linear (3) pode ser reescrito em uma forma mais simples, dada por:

$$\eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{g(y_{t-j}) - \mathbf{x}'_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \alpha_j \{g(y_{t-j}) - \eta_{t-j}\} \quad (4)$$

O modelo GARMA(p,q) fica definido por uma componente aleatória dada pela equações (1) e uma componente sistemática, ou preditor linear dada pela equação (4). Assim como nos modelos lineares generalizados, a média e variância condicionais do modelo GARMA(p,q) são dadas por

$$E(y_t|\mathbf{H}_t) = b'(\theta_t) = \mu_t, \quad (5)$$

$$Var(y_t|\mathbf{H}_t) = \varphi b''(\theta_t) = \varphi V(\mu_t). \quad (6)$$

em que $V(\mu_t)$ é chamada de função variância dependendo apenas da média μ_t .

Uma das vantagens dos modelos GARMA para séries temporais é o fato de que a componente aleatório do modelo (1) pode assumir várias distribuições, desde que esta pertença à família exponencial, o que permite que o modelo seja utilizado inclusive para dados de contagem. Na próxima seção serão apresentados alguns modelos GARMA para dados de contagem.

3 Alguns modelos GARMA

Nesta seção são apresentados três casos particulares do modelo GARMA, nos quais são consideradas as distribuições de Poisson, binomial e binomial negativa, sendo estas as distribuições discretas mais comuns.

3.1 Modelo GARMA Poisson

Supondo que Y_t é uma série de dados de contagem que pode assumir valores em $\Omega = \{0, 1, 2, \dots\}$, com distribuição de probabilidade condicional de Poisson(μ_t), que pode ser expressa como:

$$f(y_t|\mathbf{H}_t) = \exp \{y_t \log(\mu_t) - \mu_t - \log(y_t!)\} \quad (7)$$

Portanto, (7) pertence à família exponencial com $\varphi = 1$, $\theta_t = \log(\mu_t)$, $c(z, \varphi) = -\log(z!)$ e conseqüentemente $b(\theta_t) = \mu_t = \exp(\theta_t)$ e $V(\mu_t) = \mu_t$.

Para este modelo considera-se a função de ligação canônica, $g(\mu_t) = \eta_t$ dada por:

$$g(\mu_t) = \log(\mu_t). \quad (8)$$

Portanto, o modelo GARMA Poisson fica definido por (7) e (8). O preditor linear pode ser escrito como:

$$\eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{\log(y_{t-j}^*) - \mathbf{x}'_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \alpha_j \{\log(y_{t-j}^* / \mu_{t-j})\} \quad (9)$$

em que $y_{t-j}^* = \max(y_{t-j}, c)$, $0 < c < 1$.

Quando não há variáveis externas, podemos escrever $\mathbf{x}'_t \boldsymbol{\beta} = \beta_0$ e assim a equação (9) pode ser escrita como:

$$\log(\mu_t) = \beta_0 + \sum_{j=1}^p \phi_j \log(y_{t-j}^*) + \sum_{j=1}^q \alpha_j \{\log(y_{t-j}^* / \mu_{t-j})\} \quad (10)$$

3.2 Modelo GARMA Binomial

Considerando Y_t uma série de dados de contagem que pode assumir valores em $\Omega = \{0, 1, 2, \dots, M\}$ com distribuição Binomial, $B(M, \mu_t)$, tem-se que

$$f(y_t|\mathbf{H}_t) = \exp \left\{ y_t \log \left(\frac{\mu_t}{M - \mu_t} \right) + M \log \left(\frac{M - \mu_t}{M} \right) + \right. \quad (11)$$

$$\left. \log \left(\frac{\Gamma(M+1)}{\Gamma(y_t+1)\Gamma(M-y_t+1)} \right) \right\} \quad (12)$$

que pertence a família exponencial com $\varphi = 1$,

$$\theta_t = \log \left(\frac{\mu_t}{M - \mu_t} \right), \quad (13)$$

$$b(\theta_t) = -M \log \left(\frac{M - \mu_t}{M} \right), \quad (14)$$

$$c(y_t, \varphi) = \log \left(\frac{\Gamma(M + 1)}{\Gamma(y_t + 1)\Gamma(M - y_t + 1)} \right). \quad (15)$$

e função de variância é dada por $V(\mu_t) = db(\theta_t)/d\mu_t = \mu_t(m - \mu_t)/m$.

Para o modelo Binomial considera-se a função de ligação canônica, $g(\mu_t) = \eta_t$ dada por:

$$g(\mu_t) = \log \left(\frac{\mu_t}{M - \mu_t} \right) \quad (16)$$

O preditor linear é escrito na forma da equação (4) substituindo-se y_t por $y_t^* = \min\{\max(y_t, c), (M - c)\}$, com $0 < c < 1$. Se $M = 1$ tem-se o modelo GARMA(p, q) Binário. Portanto, o modelo GARMA Binomial é definido pelas equações (12) e (16).

3.3 Modelo GARMA Binomial Negativo

No modelo GARMA Binomial Negativo considera-se que Y_t é uma série de dados de contagem que pode assumir valores em $\Omega = \{0, 1, 2, \dots\}$ com distribuição Binomial Negativa, $NB(K, \mu_t)$ escrita como:

$$f(y_t | \mathbf{H}_t) = \exp \left\{ K \log \left(\frac{K}{\mu_t + K} \right) + y_t \log \left(\frac{\mu_t}{\mu_t + K} \right) + \log \left(\frac{\Gamma(K + y_t)}{\Gamma(y_t + 1)\Gamma(K)} \right) \right\}. \quad (17)$$

esta distribuição pertence a família exponencial com $\varphi = 1$,

$$\theta_t = \log \left(\frac{\mu_t}{K + \mu_t} \right), \quad (18)$$

$$b(\theta_t) = -K \log \left(\frac{K}{K + \mu_t} \right), \quad (19)$$

$$c(y_t, \varphi) = \log \left(\frac{\Gamma(K + y_t)}{\Gamma(y_t + 1)\Gamma(K)} \right). \quad (20)$$

e função de variância dada por $V(\mu_t) = \mu_t(K + \mu_t)/K$.

A função de ligação canônica, $g(\mu_t) = \eta_t$, é dada por

$$g(\mu_t) = \log \left(\frac{\mu_t}{K + \mu_t} \right). \quad (21)$$

O preditor linear é escrito na forma da equação (4) considerando $y_t^* = \max\{y_t, c\}$, com $0 < c < 1$. As equações (17) e (21) definem o modelo GARMA Binomial Negativo.

Os modelos apresentados nesta seção são apresentados de forma resumidas na Tabela 1.

Tabela 1: Modelos GARMA para dados de contagem.

	Poisson	Binomial	Binomial Negativa
Notação	$P(\mu_t)$	$B(M, \mu_t)$	$BN(K, \mu_t)$
Parâmetro de dispersão: φ	1	1	1
Função cumulante: $b(\theta_t)$	$\exp(\theta_t)$	$-M \log\left(\frac{M-\mu_t}{M}\right)$	$-K \log\left(\frac{K}{K+\mu_t}\right)$
$c(y_t, \varphi)$	$-\log(y_t!)$	$\log\left(\frac{\Gamma(M+1)}{\Gamma(y_t+1)\Gamma(M-y_t+1)}\right)$	$\log\left(\frac{\Gamma(K+y_t)}{\Gamma(y_t+1)\Gamma(K)}\right)$
$\mu_t(\theta_t)$	$\exp(\theta_t)$	$\frac{M \exp(\theta_t)}{1+\exp(\theta_t)}$	$\frac{K \exp(\theta_t)}{1-\exp(\theta_t)}$
Função de ligação (η_t)	$\log(\mu_t)$	$\log\left(\frac{\mu_t}{M-\mu_t}\right)$	$\log\left(\frac{\mu_t}{K+\mu_t}\right)$
Função de variância: $V(\mu_t)$	μ_t	$\frac{\mu_t}{M}(M-\mu_t)$	$\frac{\mu_t}{K}(K+\mu_t)$

4 Ajuste do modelo GARMA

Considerando as observações $\mathbf{H}_n = \{y_1, \dots, y_n\}$ da série temporal Y_t que satisfaça as condições dos modelos GARMA dadas em (1) e (4). Os parâmetros do modelo são $\boldsymbol{\gamma}' = (\boldsymbol{\beta}' \ \boldsymbol{\phi}' \ \boldsymbol{\alpha}')$, em que $\boldsymbol{\beta} = (\beta_0 \dots \beta_m)'$, $\boldsymbol{\phi} = (\phi_1 \dots \phi_p)'$ e $\boldsymbol{\alpha} = (\alpha_1 \dots \alpha_q)'$. Para estimar estes parâmetros vamos considerar a função de verossimilhança aproximada (condicionada nas r primeiras observações $\mathbf{H}_r = \{y_1, \dots, y_r\}$) em que $r = \max\{p, q\}$. A função de verossimilhança parcial pode ser construída considerando que Y_{t-1} e Y_t são condicionalmente independentes. Assim tem-se que

$$\begin{aligned} \mathcal{L}_a(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha} | \mathbf{H}_n) &\propto \prod_{t=r+1}^n f(y_t | \mathbf{H}_t) \\ &\propto \prod_{t=r+1}^n \exp\left\{ \frac{y_t g(\mu_t) - b(g^{-1}(\mu_t))}{\varphi} + c(y_t, \varphi) \right\}, \end{aligned} \quad (22)$$

em que $g(\mu_t)$ é a componente sistemática $g(\mu_t)$ dada por

$$g(\mu_t) = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{g(y_{t-j}^*) - \mathbf{x}'_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \alpha_j \{g(y_{t-j}^*) - g(\mu_{t-j})\} \quad (23)$$

para todo $t = r + 1, \dots, n$.

A equação (22) pode ser reescrita como

$$\mathcal{L}_a(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha} | \mathbf{H}_n) \propto \exp \left\{ \frac{1}{\varphi} \sum_{t=r+1}^n [y_t g(\mu_t) - b(g^{-1}(\mu_t))] + \sum_{t=r+1}^n c(y_t, \varphi) \right\} \quad (24)$$

e a função log-verossimilhança é dada por

$$\ell_a(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha} | \mathbf{H}_n) \propto \frac{1}{\varphi} \sum_{t=r+1}^n [y_t g(\mu_t) - b(g^{-1}(\mu_t))] + \sum_{t=r+1}^n c(y_t, \varphi). \quad (25)$$

Considerando as matrizes A e B , definidas usando-se a equação (23), dadas por

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

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

A componente sistemática (23) pode ser escrita, para as observações y_{r+1}, \dots, y_n , em forma matricial como

$$\boldsymbol{\eta} = [\mathbf{X} \quad \mathbf{A} \quad \mathbf{B}] \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\phi} \\ \boldsymbol{\alpha} \end{bmatrix},$$

em que \mathbf{X} é a matriz do modelo formada pelas linhas \mathbf{x}_t , $t = r + 1, \dots, n$.

Pode-se também denotar a componente sistemática reduzindo-a apenas a $\boldsymbol{\eta} = \mathbf{M}\boldsymbol{\gamma}$, em que $\mathbf{M} = [\mathbf{X} \quad \mathbf{A} \quad \mathbf{B}]$ é a matriz modelo de ordem $(n - r) \times (m + p + q)$.

Seja $\hat{\boldsymbol{\gamma}}$ e $\hat{\boldsymbol{\eta}} = \widehat{\mathbf{M}}\hat{\boldsymbol{\gamma}}$ os estimadores de máxima verossimilhança (EMV) de $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ respectivamente. Os EMV's $\hat{\boldsymbol{\gamma}}$ podem ser obtidos através do ajuste dos modelos GARMA dadas em (1) e (4) para y pelo método de mínimos quadrados ponderados reiterados, dado por

$$\hat{\boldsymbol{\gamma}}^{(k+1)} = \left(\widehat{\mathbf{M}}'^{(k)} \widehat{\mathbf{W}}^{(k)} \widehat{\mathbf{M}}^{(k)} \right)^{-1} \widehat{\mathbf{M}}'^{(k)} \widehat{\mathbf{W}}^{(k)} \hat{\boldsymbol{z}}^{(k)}, \quad (26)$$

com k sendo o passo da iteração, em que $\widehat{\mathbf{W}} = \text{diag}\{\widehat{w}_{r+1}, \dots, \widehat{w}_n\}$ é a matriz de pesos, em que \widehat{w}_t são os elementos $w_t = V_t^{-1} \left(\frac{dg(\mu_t)}{d\mu_t} \right)^{-2}$ avaliados fazendo-se $\mu_t = \widehat{\mu}_t$ e a variável de trabalho z_t é dada por

$$z_t = \eta_t + (y_t - \mu_t) \frac{dg(\mu_t)}{d\mu_t}. \quad (27)$$

O procedimento iterativo usando as equações (26) e (27) parte de um valor inicial $\widehat{\boldsymbol{\gamma}}^{(1)}$ e calcula $\widehat{\mathbf{M}}^{(1)}$, $\widehat{\mathbf{W}}^{(1)}$ e $\widehat{\mathbf{z}}^{(1)}$, usados para obter a próxima estimativa $\widehat{\boldsymbol{\gamma}}^{(2)}$ e $\widehat{\mathbf{z}}^{(2)}$ com (26) e (27) respectivamente. Este procedimento é realizado até que a convergência seja observada. Em Kedem & Fokianos (2002) são apresentadas as suposições necessárias para garantir que o modelo GARMA sejam bem definido e tenha estimadores consistente sob condições razoáveis.

Pode-se também fazer inferência com respeito aos parâmetros $\boldsymbol{\gamma}$ usando-se as propriedades assintóticas apresentadas em Kedem & Fokianos (2002) para $\widehat{\boldsymbol{\gamma}}$ e considerando a matriz aproximada de variância e covariâncias de $\widehat{\boldsymbol{\gamma}}$ dada por $\varphi(\widehat{\mathbf{M}}' \widehat{\mathbf{W}} \widehat{\mathbf{M}}^{-1})$.

A escolha da ordem p e q dos modelos GARMA(p, q) pode ser feita usando-se o critério BIC (*Bayesian Information Criterion*) proposto por Schwarz (1978), que é dado por

$$BIC = -2 \log\{\ell_a(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\alpha}} | \mathbf{H}_n)\} + (p + q + m) \log(n).$$

5 Previsão com os Modelos GARMA

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

$$\widehat{\eta}_t = \mathbf{x}_t \widehat{\boldsymbol{\beta}} + \sum_{j=1}^p \widehat{\phi}_j \left\{ g(y_{t-j}) - \mathbf{x}_{t-j} \widehat{\boldsymbol{\beta}} \right\} + \sum_{j=1}^q \widehat{\alpha}_j \left\{ g(y_{t-j}) - \widehat{\eta}_{t-j} \right\}, \quad (28)$$

e as médias ajustadas são obtidas através da inversa da função de ligação $\widehat{\mu}_t = g^{-1}(\widehat{\eta}_t)$, para $t = r + 1, \dots, n$.

A previsão dos valores futuros Y_{n+h} , para $h \geq 1$, supondo conhecidas as observações disponíveis até o instante $(n + 1)$

$$\mathbf{H}_{n+1} = \{x_{n+1}, x_n, \dots, x_1, y_n, y_{n-1}, \dots, y_1, \mu_n, \mu_{n-1}, \dots, \mu_1\}.$$

é dada por $\widehat{y}_{n+h} = E(Y_{n+h} | \mathbf{H}_{n+1})$. Denomina-se \widehat{y}_{n+h} por previsão de origem n e horizonte h . Nos Modelos GARMA, a previsão é feita recursivamente, de maneira simplificada, através do preditor linear para cada modelo.

Como todas as componentes do preditor são conhecidas no tempo $(n+1)$, fica fácil calcular a previsão para Y_{n+1} , tem-se

$$\hat{\eta}_{n+1} = \mathbf{x}_{n+1}\hat{\boldsymbol{\beta}} + \sum_{j=1}^p \hat{\phi}_j \left\{ g(y_{n+1-j}) - \mathbf{x}_{n+1-j}\hat{\boldsymbol{\beta}} \right\} + \sum_{j=1}^q \hat{\alpha}_j \left\{ g(y_{n+1-j}) - \hat{\eta}_{n+1-j} \right\}, \quad (29)$$

portanto, a previsão com um passo a frente ($h = 1$) é calculada por $\hat{y}_{n+1} = g^{-1}(\hat{\eta}_{n+1}) = \hat{\mu}_{n+1}$. Para $h > 1$ tem-se que

$$\hat{\eta}_{n+h} = \mathbf{x}_{n+h}\hat{\boldsymbol{\beta}} + \sum_{j=1}^p \hat{\phi}_j \left\{ g(\hat{y}_{n+h-j}) - \mathbf{x}_{n+h-j}\hat{\boldsymbol{\beta}} \right\} + \sum_{j=1}^q \hat{\alpha}_j \left\{ g(\hat{y}_{n+h-j}) - \hat{\eta}_{n+h-j} \right\}, \quad (30)$$

em que $\hat{y}_{n+h-j} = g^{-1}(\hat{\eta}_{n+h-j})$ quando $h > j$.

A variância, $\sigma_{n+h}^2 = \text{Var}(Y_{n+h}|\mathbf{H}_{n+1})$ pode ser calculado para $h = 1, 2, \dots$ usando a relação (6). Assim, tem-se que $\hat{\sigma}_{n+h}^2 = V(\hat{\mu}_{n+h})$. Intervalos de confiança para as previsões \hat{y}_{n+h} são difíceis de serem calculados uma vez que a distribuição de \hat{y}_{n+h} não são conhecidas. No entanto, fazendo uma analogia com as propriedades da distribuição Normal é possível calcular intervalos definidos por $\hat{y}_{n+h} \pm 1.96\hat{\sigma}_{n+h}$. Uma avaliação do nível de confiança deste intervalo pode se avaliado por simulação de Monte Carlo.

6 Abordagem bayesiana

Um dos objetivo da análise bayesiana dos modelos GARMA é considerar inferência sobre os parâmetros do modelo $\boldsymbol{\gamma}' = (\boldsymbol{\beta}' \boldsymbol{\phi}' \boldsymbol{\alpha}')$, em que $\boldsymbol{\beta} = (\beta_0 \dots \beta_m)'$, $\boldsymbol{\phi} = (\phi_1 \dots \phi_p)'$ e $\boldsymbol{\alpha} = (\alpha_1 \dots \alpha_q)'$ para fazer a previsão dos valores y_{t+h} contornando as dificuldades encontradas com o método clássico de máxima verossimilhança. Nesta abordagem, serão consideradas distribuições *a priori* normais, $N(\boldsymbol{\beta}_0, \sigma_\beta^2 \mathbf{I}_m)$, $N(\boldsymbol{\phi}_0, \sigma_\phi^2 \mathbf{I}_p)$ e $N(\boldsymbol{\alpha}_0, \sigma_\alpha^2 \mathbf{I}_q)$, em que os vetores de média $\boldsymbol{\beta}_0$, $\boldsymbol{\phi}_0$, $\boldsymbol{\alpha}_0$ e os parâmetros σ_β^2 , σ_ϕ^2 e σ_α^2 são hiperparâmetros conhecidos. As matrizes \mathbf{I}_m , \mathbf{I}_p e \mathbf{I}_q são matrizes identidade com dimensão escolhida adequadamente para cada vetor de parâmetro. Além disto, supondo independência entre os parâmetros, a distribuição *a priori* conjunta é dada por

$$\pi_0(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}) \propto \exp \left\{ \begin{aligned} & - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'(\sigma_\beta^2 \mathbf{I}_m)^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \\ & - \frac{1}{2}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)'(\sigma_\phi^2 \mathbf{I}_p)^{-1}(\boldsymbol{\phi} - \boldsymbol{\phi}_0) + \\ & - \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)'(\sigma_\alpha^2 \mathbf{I}_q)^{-1}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \end{aligned} \right\} \quad (31)$$

Combinando-se a verossimilhança (24) com a densidade *a priori* (31) tem-se a densidade *a posteriori* para os modelos GARMA dada por

$$\pi(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha} | \mathbf{H}_n) \propto \mathcal{L}_a(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha} | \mathbf{H}_n) \pi_0(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}) \quad (32)$$

6.1 Algoritmo MCMC

A densidade *a posteriori* dada em (32) não tem uma forma conhecida. Portanto, opta-se por um algoritmo de simulação de Monte Carlo em Cadeia de Markov (MCMC) para gerar amostras da densidade (32). O algoritmo de Metropolis-Hastings (Chib & Greenberg (1995)) é geralmente usado quando não são facilmente identificadas as distribuições condicionais a posteriori, o que impossibilita a geração direta a partir destas distribuições. A idéia do método é gerar um valor de uma distribuição auxiliar e aceitar este valor com uma certa probabilidade. O algoritmo Metropolis-Hastings é esquematizado da seguinte forma:

1. Arbitre o valor inicial para os vetores $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\phi}^{(0)}$ e $\boldsymbol{\alpha}^{(0)}$ e inicialize o contador de iterações $j = 1$.
2. Gere novos vetores de suas respectivas distribuições:

$$\boldsymbol{\beta}^{(*)} \sim N(\boldsymbol{\beta}_0, \sigma_{\boldsymbol{\beta}}^2 \mathbf{I}_m), \quad \boldsymbol{\phi}^{(*)} \sim N(\boldsymbol{\phi}_0, \sigma_{\boldsymbol{\phi}}^2 \mathbf{I}_p), \quad \boldsymbol{\alpha}^{(*)} \sim N(\boldsymbol{\alpha}_0, \sigma_{\boldsymbol{\alpha}}^2 \mathbf{I}_q)$$

3. Adote o seguinte procedimento de aceitação e rejeição dos valores gerados:
 - i. Calcule a probabilidade de aceitação para $\boldsymbol{\beta}^{(*)}$ por:

$$\varrho(\boldsymbol{\beta}^{(*)}; \boldsymbol{\beta}^{(j)}) = \min \left\{ 1, \frac{\mathcal{L}_a(\boldsymbol{\beta}^{(*)}, \boldsymbol{\phi}^{(j)}, \boldsymbol{\alpha}^{(j)} | \mathbf{H}_n)}{\mathcal{L}_a(\boldsymbol{\beta}^{(j)}, \boldsymbol{\phi}^{(j)}, \boldsymbol{\alpha}^{(j)} | \mathbf{H}_n)} \right\},$$

gere u de uma distribuição $U(0, 1)$, se $u \leq \varrho(\boldsymbol{\beta}^{(*)}; \boldsymbol{\beta}^{(j)})$ aceite o novo vetor gerado e faça $\boldsymbol{\beta}^{(j+1)} = \boldsymbol{\beta}^{(*)}$, caso contrário rejeite-o e faça $\boldsymbol{\beta}^{(j+1)} = \boldsymbol{\beta}^{(j)}$.

- ii. Calcule a probabilidade de aceitação para $\boldsymbol{\phi}^{(*)}$ por:

$$\varrho(\boldsymbol{\phi}^{(*)}; \boldsymbol{\phi}^{(j)}) = \min \left\{ 1, \frac{\mathcal{L}_a(\boldsymbol{\beta}^{(j+1)}, \boldsymbol{\phi}^{(*)}, \boldsymbol{\alpha}^{(j)} | \mathbf{H}_n)}{\mathcal{L}_a(\boldsymbol{\beta}^{(j+1)}, \boldsymbol{\phi}^{(j)}, \boldsymbol{\alpha}^{(j)} | \mathbf{H}_n)} \right\},$$

gere u de uma distribuição $U(0, 1)$, se $u \leq \varrho(\boldsymbol{\phi}^{(*)}; \boldsymbol{\phi}^{(j)})$ aceite o novo vetor gerado e faça $\boldsymbol{\phi}^{(j+1)} = \boldsymbol{\phi}^{(*)}$, caso contrário rejeite-o e faça $\boldsymbol{\phi}^{(j+1)} = \boldsymbol{\phi}^{(j)}$.

iii. Calcule a probabilidade de aceitação para $\boldsymbol{\alpha}^{(*)}$ por:

$$\varrho(\boldsymbol{\alpha}^{(*)}; \boldsymbol{\alpha}^{(j)}) = \min \left\{ 1, \frac{\mathcal{L}_a(\boldsymbol{\beta}^{(j+1)}, \boldsymbol{\phi}^{(j+1)}, \boldsymbol{\alpha}^{(*)} | \mathbf{H}_n)}{\mathcal{L}_a(\boldsymbol{\beta}^{(j+1)}, \boldsymbol{\phi}^{(j+1)}, \boldsymbol{\alpha}^{(j)} | \mathbf{H}_n)} \right\},$$

gere u de uma distribuição $U(0, 1)$, se $u \leq \varrho(\boldsymbol{\alpha}^{(*)}; \boldsymbol{\alpha}^{(j)})$ aceite o novo vetor gerado e faça $\boldsymbol{\alpha}^{(j+1)} = \boldsymbol{\alpha}^{(*)}$, caso contrário rejeite-o valor gerado e faça $\boldsymbol{\alpha}^{(j+1)} = \boldsymbol{\alpha}^{(j)}$.

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

O procedimento de geração de uma amostra da distribuição de equilíbrio (32) descarta uma sequência gerada inicialmente (período de aquecimento) necessário para eliminar o efeito das condições iniciais, arbitradas, na convergência da cadeia para à distribuição de equilíbrio e faz uma reamostragem da amostra restante para diminuir a correlação, provocada pelo procedimento de aceitação e rejeição dos valores gerados, dentro das cadeias geradas. Para monitoramento do período de aquecimento e estacionariedade das cadeias geradas foi utilizado o critério proposto por Geweke (1992). Denotando-se por $\boldsymbol{\gamma}^{(j)} = (\boldsymbol{\beta}^{(j)'} \boldsymbol{\phi}^{(j)'} \boldsymbol{\alpha}^{(j)'})$, em que $\boldsymbol{\beta}^{(j)} = (\beta_0^{(j)} \dots \beta_m^{(j)})'$, $\boldsymbol{\phi}^{(j)} = (\phi_1^{(j)} \dots \phi_p^{(j)})'$ e $\boldsymbol{\alpha}^{(j)} = (\alpha_1^{(j)} \dots \alpha_q^{(j)})'$, com $j = 1, \dots, Q$ a amostra gerada, resultante, dos parâmetros do modelo. Então, sumários das densidades *a posteriori* marginais são apresentados com estimativas de Monte Carlo calculadas com estas amostras geradas.

7 Critério de seleção de modelo

Para selecionar o melhor modelo $\text{GARMA}(p, q)$, ou a melhor ordem p e q de um modelo GARMA , a ser ajustado a uma série temporal $\{y_t, t = 1, 2, \dots\}$ de dados de contagem, utiliza-se, neste trabalho, alguns critérios Bayesianos de seleção de modelo.

7.1 Critério de Informação Bayesiano (BIC)

O critério de informação Bayesiano (ou *Bayesian Information Criterion*, BIC) é um critério de seleção de modelo proposto por Schwarz (1978) e modificado por Carlin & Louis (2000) para ser utilizado no contexto da inferência Bayesianas levando em conta a densidade a posteriori dos parâmetros do modelo ajustado. Este critério, denotado por $EBIC = E\{BIC\}$, faz uma ponderação entre o valor esperado, *a posteriori*, do máximo da função log-verossimilhança e o número de parâmetros do modelo ajustado. Considera-se

como o modelo que melhor se ajusta, aquele que apresenta o menor valor do $EBIC$ dado por

$$EBIC = -2E\{\ell_a(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}|\mathbf{H}_n)\} + (p + q + m) \log(n). \quad (33)$$

em que $E\{\ell_a(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}|\mathbf{H}_n)\}$ é o valor esperado, tomado com relação à densidade *a posteriori*, da máxima log-verossimilhança e n é o tamanho da série usada no ajuste do modelo GARMA(p,q).

7.2 Critério de Informação Desvio (DIC)

O critério de informação desvio (ou *Deviance Information Criterion*, DIC) é muito usado no método Bayesiano de seleção de modelo quando se dispõe de amostras geradas da distribuição *a posteriori* dos parâmetros dos modelos. O critério DIC, introduzido por Spiegelhalter *et al.* (2002), é definido a partir da função D , dada por

$$D(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}) = -2\ell_a(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}|\mathbf{H}_n) \quad (34)$$

O desvio, DIC, é dado por

$$DIC = 2\bar{D} - \hat{D} \quad (35)$$

em que $\bar{D} = E\{D(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha})\}$ é a média *a posteriori* de D e \hat{D} é o valor de D calculado com valores médios *a posteriori* dos parâmetros $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\alpha}}$.

Ambos \bar{D} e \hat{D} são calculados com a amostra $\boldsymbol{\gamma}^{(j)} = (\boldsymbol{\beta}^{(j)'} \boldsymbol{\phi}^{(j)'} \boldsymbol{\alpha}^{(j)'})$, gerada da densidade *a posteriori* como

$$\begin{aligned} \bar{D} &= \frac{1}{Q} \sum_{j=1}^Q D(\boldsymbol{\gamma}^{(j)}) \\ \hat{D} &= -2\ell_a(\hat{\boldsymbol{\gamma}}) \end{aligned}$$

em que

$$\hat{\boldsymbol{\gamma}} = \frac{1}{Q} \left\{ \sum_{j=1}^Q \boldsymbol{\gamma}^{(j)} \right\}.$$

Valores menores para DIC indicam os melhores modelos, inclusive esses valores podem ser negativos.

7.3 Densidade Condicional Preditiva Ordenada (CPO)

A densidade condicional preditiva ordenada (CPO), foi inicialmente proposta por Geisser & Eddy (1979) e foi discutida em Gelfand *et al.* (1992). Para calcular esta densidade preditiva ordenada associada a um valor da série y_t , denotada por CPO_t , vamos considerar a densidade *a posteriori* $\pi(\boldsymbol{\gamma}|\mathbf{H}_t)$, dada na equação (32), condicionada nas informações disponíveis até o tempo t , denotada aqui por

$$\mathbf{H}_t = \{x_t, x_{t-1}, \dots, x_1, y_{t-1}, y_{t-2}, \dots, y_1, \mu_{t-1}, \mu_{t-2}, \dots, \mu_1\}$$

e a densidade da observação y_t , denotada por $f(y_t|\boldsymbol{\gamma}, \mathbf{H}_t)$, dada na equação (1). Então tem-se que

$$CPO_t = \int f(y_t|\boldsymbol{\gamma}, \mathbf{H}_t)\pi(\boldsymbol{\gamma}|\mathbf{H}_t)d\boldsymbol{\gamma} \quad (36)$$

usando a equação (32), a equação (36) pode ser escrita como

$$CPO_t = \int f(y_t|\boldsymbol{\gamma}, \mathbf{H}_t) \left\{ \frac{\mathcal{L}_a(\boldsymbol{\gamma}|\mathbf{H}_t)\pi_0(\boldsymbol{\gamma})}{\int \mathcal{L}_a(\boldsymbol{\gamma}|\mathbf{H}_t)\pi_0(\boldsymbol{\gamma})d\boldsymbol{\gamma}} \right\} d\boldsymbol{\gamma} \quad (37)$$

sabendo-se que $f(y_t|\boldsymbol{\gamma}, \mathbf{H}_t)\mathcal{L}_a(\boldsymbol{\gamma}|\mathbf{H}_t) = \mathcal{L}_a(\boldsymbol{\gamma}|\mathbf{H}_{t+1})$, em que

$$\mathbf{H}_{t+1} = \{x_{t+1}, x_t, \dots, x_1, y_t, y_{t-1}, \dots, y_1, \mu_t, \mu_{t-1}, \dots, \mu_1\}.$$

A equação (37) pode ser escrita como

$$\begin{aligned} CPO_t &= \int \frac{\mathcal{L}_a(\boldsymbol{\gamma}|\mathbf{H}_{t+1})\pi_0(\boldsymbol{\gamma})}{\left\{ \int \frac{1}{f(y_t|\boldsymbol{\gamma}, \mathbf{H}_t)} \mathcal{L}_a(\boldsymbol{\gamma}|\mathbf{H}_{t+1})\pi_0(\boldsymbol{\gamma})d\boldsymbol{\gamma} \right\}} d\boldsymbol{\gamma} \\ &= \frac{\int \mathcal{L}_a(\boldsymbol{\gamma}|\mathbf{H}_{t+1})\pi_0(\boldsymbol{\gamma})d\boldsymbol{\gamma}}{\int \frac{1}{f(y_t|\boldsymbol{\gamma}, \mathbf{H}_t)} \mathcal{L}_a(\boldsymbol{\gamma}|\mathbf{H}_{t+1})\pi_0(\boldsymbol{\gamma})d\boldsymbol{\gamma}} \end{aligned} \quad (38)$$

Após alguma algebra mostra-se que

$$CPO_t = \left\{ \int \frac{1}{f(y_t|\boldsymbol{\gamma}, \mathbf{H}_t)} \pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1})d\boldsymbol{\gamma} \right\}^{-1}. \quad (39)$$

Portanto, a densidade preditiva ordenada, CPO_t , pode ser estimada usando-se uma amostra $\{\boldsymbol{\gamma}^{(j)}, j = 1, \dots, Q\}$, gerada com o algoritmo MCMC, da densidade *a posteriori* $\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1})$. O estimador de Monte Carlo da CPO_t , é dado por:

$$\widehat{CPO}_t = \left[\frac{1}{Q} \sum_{j=1}^Q \frac{1}{f(y_t|\boldsymbol{\gamma}^{(j)}, \mathbf{H}_t)} \right]^{-1}. \quad (40)$$

As estimativas de \widehat{CPO}_t , para $t = r + 1, \dots, n$, podem ser usadas para selecionar o melhor modelo GARMA(p, q) que se ajusta a uma série de dados de contagem. Seleciona-se como melhor modelo, aquele que apresentar maior valor do $\widehat{CPO} = \prod_{t=r+1}^n \widehat{CPO}_t$. Costuma-se utilizar o critério definido por $\log(\widehat{CPO}) = \sum_{t=r+1}^n \log(\widehat{CPO}_t)$.

8 Identificação de pontos influentes

Uma observação y_t pode ser identificada como influente, avaliando-se divergência de Kullback-Leibler provocada por esta observação na entre estas duas densidades *a posteriori* obtidas com e sem esta observação. Esta medida, denotada por $KL(\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1}); \pi(\boldsymbol{\gamma}|\mathbf{H}_t))$, é avaliada calculando-se a divergência de Kullback-Leibler entre as densidades $\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1})$, obtida com todas as observações até y_t , e $\pi(\boldsymbol{\gamma}|\mathbf{H}_t)$ que representa a densidade *a posteriori* obtida ao se retirar a observação y_t da amostra. A divergência de Kullback-Leibler é dada por

$$KL(\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1}); \pi(\boldsymbol{\gamma}|\mathbf{H}_t)) = \int \pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1}) \log \left(\frac{\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1})}{\pi(\boldsymbol{\gamma}|\mathbf{H}_t)} \right) d\boldsymbol{\gamma}. \quad (41)$$

Em Cho *et al.* (2009) foi mostrado que (41) pode ser avaliada considerando-se a CPO_t dada em (39), como

$$KL(\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1}); \pi(\boldsymbol{\gamma}|\mathbf{H}_t)) = \int \log(f(y_t|\mathbf{H}_t))\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1})d\boldsymbol{\gamma} - \log(CPO_t). \quad (42)$$

A divergência (42), pode ser estimada usando-se uma amostra $\{\boldsymbol{\gamma}^{(j)}, j = 1, \dots, M\}$, gerada com o algoritmo MCMC, da densidade *a posteriori* $\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1})$. O estimador de Monte Carlo de $KL(\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1}); \pi(\boldsymbol{\gamma}|\mathbf{H}_t))$, é dado por

$$\widehat{KL}(\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1}); \pi(\boldsymbol{\gamma}|\mathbf{H}_t)) = \frac{1}{Q} \sum_{j=1}^Q \log(f(y_t|\boldsymbol{\gamma}^{(j)}, \mathbf{H}_t)) - \log(\widehat{CPO}_t), \quad (43)$$

em que \widehat{CPO}_t é a estimativa de Monte Carlos da CPO_t , dada na equação (40).

Uma medida de calibração para a divergência $KL(\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1}); \pi(\boldsymbol{\gamma}|\mathbf{H}_t))$, proposto por McCulloch (1989), denotada aqui por ρ_t , é obtida resolvendo-se a equação $KL(\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1}); \pi(\boldsymbol{\gamma}|\mathbf{H}_t)) = K(B(0.5), B(\rho_t)) = -\log 4\rho_t(1 - \rho_t)/2$, em que $B(\rho_t)$ denota a distribuição de Bernoulli com probabilidade de sucesso ρ_t . Isto significa que descrever os resultados com a densidade *a posteriori* completa, $\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1})$, em vez da densidade *a posteriori* obtida após

removida a observação y_t , $\pi(\boldsymbol{\gamma}|\mathbf{H}_t)$, é equivalente a descrever um evento não observado com tendo probabilidade ρ_t , quando a probabilidade correta é 0.5. Resolvendo-se a equação para ρ_t , tem-se que

$$\rho_t = \frac{1}{2} \left\{ 1 + \sqrt{1 - \exp[-2KL(\pi(\boldsymbol{\gamma}|\mathbf{H}_{t+1}); \pi(\boldsymbol{\gamma}|\mathbf{H}_t))]} \right\}.$$

Isto implica que $0.5 \leq \rho_t \leq 1$. Para $\rho_t \gg 0.5$ tem-se que a observação y_t é uma observação influente.

9 Previsão Bayesiana

Na abordagem Bayesiana, o preditor linear η_t , dado na equação (4), além de depender das observações \mathbf{H}_t , também depende do vetor de parâmetros $\boldsymbol{\gamma}' = (\boldsymbol{\beta}' \ \boldsymbol{\phi}' \ \boldsymbol{\alpha}')$, os quais são variáveis aleatórias com densidade *a posteriori* dada por (32). Neste contexto, o preditor linear (4) é denotado como $\eta_t(\boldsymbol{\gamma}) = E(Y_t|\boldsymbol{\gamma}, \mathbf{H}_t)$, para $t = r + 1, \dots, n$.

O preditor linear *a posteriori* $\eta_t = E(Y_t|\mathbf{H}_t)$ é dado por

$$\eta_t = \int \eta_t(\boldsymbol{\gamma})\pi(\boldsymbol{\gamma}|\mathbf{H}_t)d\boldsymbol{\gamma}. \quad (44)$$

em que $\eta_t(\boldsymbol{\gamma})$ é dado por

$$\eta_t(\boldsymbol{\gamma}) = \mathbf{x}_t\boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{g(y_{t-j}) - \mathbf{x}_{t-j}\boldsymbol{\beta}\} + \sum_{j=1}^q \alpha_j \{g(y_{t-j}) - \eta_{t-j}(\boldsymbol{\gamma})\}, \quad (45)$$

Portanto, o estimador de Monte Carlo do preditor linear η_t , para uma amostra $\{\boldsymbol{\gamma}^{(j)}, j = 1, \dots, M\}$, gerada com o algoritmo MCMC, da densidade *a posteriori* $\pi(\boldsymbol{\gamma}|\mathbf{H}_t)$, é dado por:

$$\hat{\eta}_t = \frac{1}{M} \sum_{j=1}^M \eta_t(\boldsymbol{\gamma}^{(j)}). \quad (46)$$

As médias $\mu_t(\boldsymbol{\gamma})$, para $t = r + 1, \dots, n$, são obtidas com a inversa da função de ligação $\mu_t(\boldsymbol{\gamma}) = g^{-1}(\eta_t(\boldsymbol{\gamma}))$ e o estimador de Monte Carlo para μ_t é dada por

$$\hat{\mu}_t = \frac{1}{M} \sum_{j=1}^M g^{-1}(\eta_t(\boldsymbol{\gamma}^{(j)})). \quad (47)$$

Para prever valores futuros Y_{n+h} da série de dados de contagem, para $h \geq 1$ e supondo conhecidas as observações disponíveis até o tempo $(n + 1)$, dada por

$$\mathbf{H}_{n+1} = \{x_{n+1}, x_n, \dots, x_1, y_n, y_{n-1}, \dots, y_1, \mu_n, \mu_{n-1}, \dots, \mu_1\}$$

é preciso calcular o valor esperado condicional $\hat{y}_{n+h} = E(Y_{n+h}|\mathbf{H}_{n+1})$, no contexto Bayesiano. Como todas as componentes do preditor $\eta_{n+1}(\boldsymbol{\gamma})$ são conhecidas, fica fácil calcular a previsão com um passo a frente.

Para $h = 1$, tem-se que

$$\eta_{n+1}(\boldsymbol{\gamma}) = \mathbf{x}_{n+1}\boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{g(y_{n+1-j}) - \mathbf{x}_{n+1-j}\boldsymbol{\beta}\} + \sum_{j=1}^q \alpha_j \{g(y_{n+1-j}) - \eta_{n+1-j}(\boldsymbol{\gamma})\}, \quad (48)$$

Portanto, estimadores de Monte Carlo para o preditor linear, η_{n+1} , e para a previsão, μ_{n+1} , são calculados substituindo-se a equação (48) nas equações (46) e (47) respectivamente. A previsão com um passo a frente ($h = 1$) é dada por $\hat{y}_{n+1} = \hat{\mu}_{n+1}$.

Para $h > 1$ tem-se que

$$\eta_{n+h}(\boldsymbol{\gamma}) = \mathbf{x}_{n+h}\boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{g(\hat{y}_{n+h-j}) - \mathbf{x}_{n+h-j}\boldsymbol{\beta}\} + \sum_{j=1}^q \alpha_j \{g(\hat{y}_{n+h-j}) - \eta_{n+h-j}(\boldsymbol{\gamma})\}, \quad (49)$$

em que usa-se $\hat{y}_{n+h-j} = \hat{\mu}_{n+h-j}$ quando $h > j$.

Estimadores de Monte Carlo para o preditor linear, η_{n+h} , e para a previsão, μ_{n+h} , são calculados substituindo-se a equação (49) nas equações (46) e (47) respectivamente. A previsão para Y_{n+h} com $h > 1$ é dada por $\hat{y}_{n+h} = \hat{\mu}_{n+h}$. Intervalos de credibilidade para $\hat{\mu}_{n+h}$ podem ser calculados estimando os percentis, $100\delta\%$, $100(1 - \delta)\%$ da amostra $\{\mu_{n+h}^{(j)} = g^{-1}(\eta_{n+h}(\boldsymbol{\gamma}^{(j)})), j = 1, \dots, M\}$ obtidas com a amostra gerada via MCMC. A variância, $\sigma_{n+h}^2 = \text{Var}(Y_{n+h}|\mathbf{H}_{n+1})$ pode ser calculado para $h = 1, 2, \dots$ usando a relação (6). Assim, tem-se que $\sigma_{n+h}^2(\boldsymbol{\gamma}) = V(\mu_{n+h}(\boldsymbol{\gamma}))$.

10 Aplicações

Para ilustrar a metodologia Bayesiana proposta no ajuste dos modelos e avaliar o desempenho dos critérios de seleção na escolha do melhor modelo GARMA(p, q), foi feito um estudo com séries simuladas e uma análise de uma série real de incidência de Dengue na Paraíba.

10.1 estudo com séries simuladas

Neste estudo foram geradas séries de tamanho $n = 100$ com determinado modelo GARMA, considerando cada uma das distribuições abordadas neste trabalho (Poisson, Binomial, Binomial Negativo) e para cada uma destas séries foram ajustados modelos GARMA com diferentes ordens (p, q) . Neste estudo foi adotado $m = 1$, ou seja, $\boldsymbol{\beta} = \beta_0$. Para selecionar a ordem (p, q) do modelo GARMA que melhor se ajusta as séries geradas foram calculados os critérios de seleção *EBIC*, *DIC* e a *LogCPO*, apresentados na Seção 7.

No ajuste de cada modelo considerou-se, no algoritmo Metropolis-Hastings uma cadeia de tamanho 10.0000 e foram descartados como período de aquecimento 50% da cadeia gerada, para diminuir a correlação dentro de cada cadeia, devido ao processo de aceitação-rejeição, foi feita uma reamostragem a cada 5 valores gerados, resultando em uma amostra de tamanho 1000 com as quais foram calculadas estimativas de Monte Carlo para se obter um sumário da densidade a posteriori marginal de cada parâmetro dos modelos.

A Tabela 2 apresenta os critérios de seleção calculados para os modelos GARMA(p, q), com $p = 1, 2$ e $q = 0, 1, 2$, ajustados à série gerada com o modelo GARMA(2,1), com parâmetros $\beta_0 = 1$, $\phi_1 = 0.5$, $\phi_2 = -0.6$, $\alpha_1 = -0.8$ e com a distribuição de Poisson.

Tabela 2: Critérios de seleção para os modelos GARMA(p, q) Poisson

Modelos	EBIC	DIC	LogCPO
GARMA(1,0)	1401	1390	-695
GARMA(2,0)	518	499	-250
GARMA(1,1)	844	831	-418
GARMA(2,1)	382	365	-183
GARMA(2,2)	488	465	-232

Observa-se na Tabela 2 que os critérios utilizados indicam a ordem correta do modelo que melhor se ajusta a série gerada. Um sumário *a posteriori* do modelo ajustado é apresentado na Tabela 3 para que se possa comparar a qualidade do ajuste com o verdadeiro modelo que foi usado para gerar a série.

Tabela 3: Sumário *a posteriori* do modelo GARMA(2,1) Poisson

Parâmetros	Real	Média	Mediana	Moda	I.C.(95%)
β_0	1.0	1.004	1.003	1.005	(0.987, 1.019)
ϕ_1	0.5	0.502	0.502	0.510	(0.483, 0.517)
ϕ_2	-0.6	-0.597	-0.596	-0.588	(-0.614,-0.578)
α_1	-0.8	-0.772	-0.773	-0.748	(-0.841,-0.702)

Resultados semelhantes aos dos modelos de Poisson, foram obtidos com os modelos Binomial. Na Tabela 4 são vistos os critérios de seleção calculados para os modelos $\text{GARMA}(p,q)$, com $p = 1, 2$ e $q = 0, 1, 2$, ajustados à série gerada com o modelo $\text{GARMA}(2,1)$, com parâmetros $\beta_0 = 0.5$, $\phi_1 = 0.5$, $\phi_2 = -0.6$, $\alpha_1 = -0.8$ e distribuição Binomial(M, μ_t), com $M = 20$. Um sumário *a posteriori* do modelo GARMA Binomial, ajustado é apresentado na Tabela 5.

Tabela 4: Critérios de seleção para os modelos $\text{GARMA}(p,q)$ Binomial

Modelos	EBIC	DIC	LogCPO
$\text{GARMA}(1,0)$	1252	1238	-619
$\text{GARMA}(2,0)$	793	774	-387
$\text{GARMA}(1,1)$	376	322	-181
$\text{GARMA}(2,1)$	349	315	-166
$\text{GARMA}(2,2)$	368	317	-173

Tabela 5: Sumário *a posteriori* do modelo $\text{GARMA}(2,1)$ Binomial

Parâmetros	Real	Média	Mediana	Moda	I.C.(95%)
β_0	0.5	0.502	0.502	0.519	(0.482, 0.519)
ϕ_1	0.5	0.501	0.502	0.488	(0.483, 0.518)
ϕ_2	-0.6	-0.601	-0.601	-0.622	(-0.619,-0.518)
α_1	-0.8	-0.795	-0.795	-0.796	(-0.807,-0.783)

Para avaliar o desempenho dos critérios de seleção na escolha da ordem do modelo $\text{GARMA}(p,q)$ ajustado com a distribuição Binomial Negativa, consideramos uma série de tamanho $n = 100$ gerada com o modelo $\text{GARMA}(2,2)$ com a distribuição Binomial negativa $BN(K, \mu_t)$, com $K = 50$ e para a esta séries foram ajustados modelos $\text{GARMA}(p,q)$, com $p = 1, 2$ e $q = 0, 1, 2$. Os valores de cada critério são apresentados na Tabela 6 na qual destaca-se o bom desempenho dos critérios na escolha da ordem correta do modelo. Um sumário *a posteriori* do modelo $\text{GARMA}(2,1)$ ajustado com a distribuição Binomial Negativa é apresentada na Tabela 7.

Para finalizar este estudo com dados simulados, consideramos as três séries geradas com cada uma das distribuições, Poisson, Binomial e Binomial Negativa e ajustamos a cada uma destas séries um modelo $\text{GARMA}(p,q)$ com as três distribuições. Os modelos que melhor se ajustou a cada série foi selecionado e o cálculo dos critérios são apresentados na Tabela 8.

Tabela 6: Critérios de seleção para os modelos GARMA(p,q) Binomial Negativa

Modelos	EBIC	DIC	LogCPO
GARMA(1,0)	649	635	-318
GARMA(2,0)	669	651	-326
GARMA(1,1)	406	394	-197
GARMA(2,1)	398	381	-190
GARMA(2,2)	484	462	-232

Tabela 7: Sumário *a posteriori* do modelo GARMA(2,1) Binomial Negativa

Parâmetros	Real	Média	Mediana	Moda	I.C.(95%)
β_0	0.5	0.499	0.499	0.487	(0.481, 0.518)
ϕ_1	0.3	0.298	0.298	0.286	(0.279, 0.317)
ϕ_2	-0.6	-0.601	-0.602	-0.620	(-0.621,-0.580)
α_1	-0.8	-0.797	-0.798	-0.785	(-0.806,-0.788)

Tabela 8: Critérios de Seleção para escolha do melhor modelo ajustado

Modelos Ajustados	Critérios	Modelos Verdadeiros		
		Poisson	Binomial	B. Negativa
Poisson	EBIC	382	515	422
	DIC	365	497	406
	LogCPO	-183	-248	-204
Binomial	EBIC	418	349	428
	DIC	381	315	412
	LogCPO	-201	-166	-207
B. Negativa	EBIC	626	524	398
	DIC	604	502	381
	LogCPO	-303	-251	-190

10.2 Análise da série de incidência de Dengue na Paraíba

Nesta seção é analisada a série de incidência de casos de Dengue na Paraíba (número de notificações/100 habitantes), no período de janeiro de 1998 a dezembro de 2005. Para esta série foram ajustados modelos GARMA(1,1) com as distribuições de Poisson, Binomial e Binomial Negativa. Os critérios de seleção do melhor modelo são apresentados na Tabela 9.

Os critérios indicam que os modelos ajustados com as distribuições Poisson e Binomial Negativa praticamente ajustam-se igualmente bem à série analisada. A opção pelo modelo GARMA-Poisson é feita por questão de

Tabela 9: Critérios para seleção do melhor modelo ajustado à série de incidência de Dengue na Paraíba

Modelos Ajustados	Critérios de Seleção		
	EBIC	DIC	LogCPO
Poisson	468	456	-229
Binomial	519	507	-257
B. Negativa	469	457	-230

simplicidade. Um sumário da densidade *a posteriori* dos parâmetros deste modelo é apresentado na Tabela 10.

Tabela 10: Sumário da distribuição a posteriori do modelo GARMA(1,1) com distribuição de Poisson ajustada a incidência de Dengue na Paraíba

Parâmetros	Média	Mediana	Moda	IC 95%
β_0	0.5405	0.5406	0.4191	(0.4040,0.7003)
ϕ_1	0.7654	0.7661	0.8529	(0.7023,0.8244)
α_1	0.4932	0.4942	0.4615	(0.4114,0.5688)

Na Figura 1 são apresentados os histogramas das amostras gerada distribuição *a posteriori* de cada parâmetro.

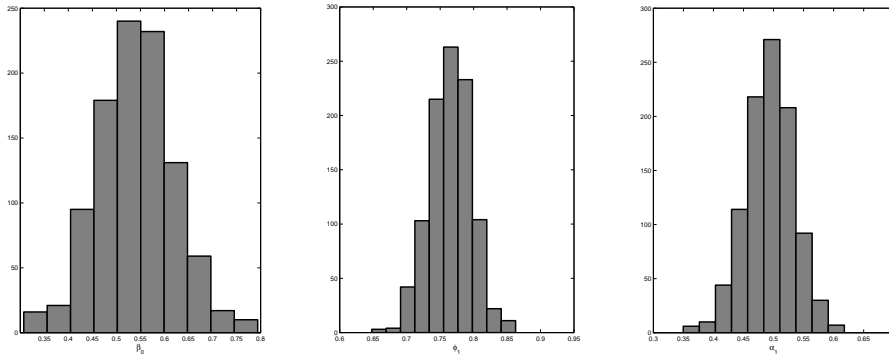


Figura 1: Histogramas das amostras geradas da distribuição *a posteriori* dos parâmetros do modelo GARMA(1,1) Poisson

A influência de cada ponto da série, avaliada com a divergência de Kullback-Leibler, calibrada, é apresentada na Figura 2. Nesta figura, adotando-se como limite de calibração $\rho_t > 0.90$, observa-se que somente a observação correspondente a $t = 50$ apresenta-se como uma medida influente.

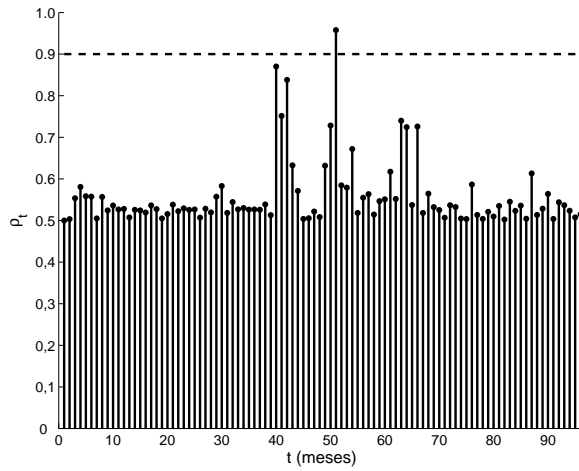


Figura 2: Medida de influência das observações com o modelo GARMA(1,1) de Poisson

Na Figura 3 são apresentadas as séries real e a série das médias $\hat{\mu}_t$ ajustadas.

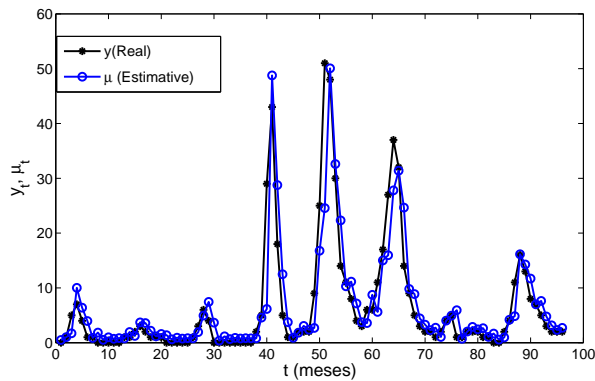


Figura 3: Série de incidência de Dengue na Paraíba e médias ($\hat{\mu}_t$) ajustada com o modelo GARMA(1,1) de Poisson

11 Conclusão

Neste trabalho foi proposta uma abordagem Bayesiana para ajuste de modelos GARMA(p,q) para séries de dados de contagem, considerando as distribuições de Poisson, Binomial e Binomial Negativa. Foram considerados

também os critérios de seleção de modelos BIC, DIC e a densidade condicional preditiva ordenada (CPO). Foi considerada também uma medida de divergência calibrada baseada na divergência de Kullback-Leibler para avaliar uma medida da influência de cada observação da série.

Foram realizados estudos com séries simuladas para verificar o desempenho dos critérios considerados na escolha da ordem (p,q) dos modelos GARMA e na seleção do melhor modelo (Poisson, Binomial ou Binomial Negativo) para representar uma série de dados de contagem. Este estudo revelou que os critérios apresentaram bom desempenho na seleção de ambas a ordem e a distribuição do modelo que melhor se ajusta a uma dada série de contagem.

A abordagem Bayesiana proposta neste trabalho foi aplicada na análise de uma série real, de incidência de Dengue (número de notificações/100 habitantes), no estado da Paraíba. O modelo GARMA(1,1) com a distribuição de Poisson foi o modelo que melhor se ajustou a série analisada. A medida de divergência calibrada apontou a observação feita na semana $t = 50$ que corresponde ao mês de fevereiro de 2002 como uma observação influente no ajuste.

Por fim, concluí-se que a abordagem Bayesiana proposta mostrou-se viável e eficiente para o ajuste dos modelos GARMA para séries de contagem e os critérios propostos mostram-se adequados para a seleção do melhor modelo que se ajusta a uma série de dados de contagem.

Referências

- Al-Osh, M. A. & Alzaid, A. A. (1987). First-order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, **8**, 261–275.
- Benjamin, M. A., Rigby, R. A. & Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, **98**, 214–223.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**(3), 307–327.
- Carlin, B. P. & Louis, T. A. (2000). *Bayes and Empirical Bayes methods for data analysis*. Chapman and Hall/CRC Press, USA, second edition.
- Chib, S. & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *American Statistician*, pages 327–335.

- Cho, H., Ibrahim, J. G., Sinha, D. & Zhu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics*, **65**, 116–124.
- Cox, R. D. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, **8**, 93–115.
- Davis, R. & Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika*, **96**, 735–749.
- Davis, R., Dunsmuir, W. T. M. & Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, **87**(3), 491–505.
- Davis, R., Dunsmuir, W. T. M. & Streett, S. B. (2003). Observation-driven models for Poisson counts. *Methodol. Comput. Appl. Probab*, **90**(4), 777–790.
- Davis, R., Dunsmuir, W. T. M. & Streett, S. B. (2005). Maximum likelihood estimation for an observation driven model for poisson counts. *Methodol. Comput. Appl. Probab.*, **7**, 149–159.
- Davis, R. A., Dunsmuir, W. T. M. & Wang, Y. (1999). Modelling time series of count data. In S. Ghosh, editor, *Asymptotics, Nonparametric & Time Series*, pages 63–114, New York. Marcel Dekker.
- Du, J. & Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, **12**, 129–142.
- Ferland, R., Latour, A. & Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, **27**, 923–942.
- Fokianos, K. & Kedem, B. (2004). Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis*, **25**(2), 173–197.
- Fokianos, K., Rahbek, A. & Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, **104**(488), 1430–1439.
- Freeland, R. K. & McCabe, B. P. M. (2004). Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, **25**(5), 701–722.
- Geisser, S. & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, **74**(365), 153–160.

- Gelfand, A. E., Dey, D. K. & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith, editors, *Bayesian Statistics*, volume 4, pages 147–167, New York. Oxford University Press.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith, editors, *Bayesian Statistics*, volume 4, pages 169–193, New York. Oxford University Press.
- Heinen, A. (2003). Modelling time series count data: An autoregressive conditional poisson model. *Available at SSRN: <http://ssrn.com/abstract=1117187>*.
- Jung, R. C. & Tremayne, A. R. (2011). Useful models for time series of counts or simply wrong ones? *Advances in Statistical Analysis*, **95**, 59–91.
- Jung, R. C., Kukuk, M. & Liesenfeld, R. (2006). Time series of count data: modeling, estimation and diagnostics. *Computational Statistics & Data Analysis*, **51**, 2350–2364.
- Kedem, B. & Fokianos, K. (2002). *Regression Models for Time Series Analysis*. John Wiley & Sons, New Jersey, USA.
- Li, W. K. L. (1994). Times series model based on generalized linear models: Some further results. *Biometrics*, **50**, 506–511.
- Liesenfeld, R., Nolte, I. & Pohlmeier, W. (2006). Modelling financial transaction price movements: a dynamic integer count data model. *Empirical Econom.*, **30**, 795–825.
- McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*. Monographs on Statistics and Applied Probability. London: Chapman and Hall, second edition.
- McCulloch, R. E. (1989). Local model influence. *Journal of the American Statistical Association*, **84**(406), 473–478.
- McKenzie, E. (1988). Some ARMA models for dependent sequences of poisson counts. *Advances in Applied Probability*, **20**, 822–835.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, **6**(2), 461–464.

- SHEPHARD, N. (1995). Generalized linear autoregressions. (Unpublished paper, Oxford University, UK).
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, **64**(4), 583–639.
- Wei, C. H. (2007). Controlling correlated processes of Poisson counts. *Qual Reliab Eng Int*, **23**(6), 741–754.
- Wei, C. H. (2008). Thinning operations for modelling time series of counts - a survey. *Advances in Statistical Analysis*, **92**, 319–341.
- Wei, C. H. (2009). Modelling time series of counts with overdispersion. *Statistical Methods & Applications*, **18**, 507–519.
- Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, **75**, 621–629.
- Zhu, F. (2011). A negative binomial interger-valued GARCH model. *Journal of time series analysis*, **32**, 54–67.