



**LUIZ OTÁVIO DE OLIVEIRA PALA**

**UM ESTUDO DE MODELOS PARA SÉRIES TEMPORAIS DE  
CONTAGEM**

**LAVRAS – MG**

**2023**

**LUIZ OTÁVIO DE OLIVEIRA PALA**

**UM ESTUDO DE MODELOS PARA SÉRIES TEMPORAIS DE CONTAGEM**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Profa. Dra. Thelma Sáfyadi  
Orientadora

**LAVRAS – MG**  
**2023**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Pala, Luiz Otávio de Oliveira

Um estudo de modelos para séries temporais de contagem /  
Luiz Otávio de Oliveira Pala. – 2023.

112 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2023.

Orientador: Profa. Dra. Thelma Sáfyadi.

Coorientador: .

Bibliografia.

1. GARMA. 2. Inferência Bayesiana. 3. MCMC. I. Sáfyadi,  
Thelma. II. Título.

**LUIZ OTÁVIO DE OLIVEIRA PALA**

**UM ESTUDO DE MODELOS PARA SÉRIES TEMPORAIS DE CONTAGEM  
A STUDY OF COUNT TIME SERIES MODELS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 27 de abril de 2023.

Profa. Dra. Gislene Araujo Pereira	UNIFAL
Prof. Dr. Luiz Ricardo Nakamura	UFLA
Prof. Dr. Manoel Vítor de Souza Veloso	UNIFAL
Prof. Dr. Paulo Henrique Sales Guimarães	UFLA

Profa. Dra. Thelma Sáfydi  
Orientadora

**LAVRAS – MG  
2023**

## **AGRADECIMENTOS**

Inicio os meus agradecimentos aos meus familiares e amigos. Em especial, aos meus pais, Maria e Luiz e ao meu irmão Mateus, que me apoiam e incentivam em todas as escolhas.

Aos meus amigos Aline, Ana Flávia, Bruna, Daiane, Isabella, Jéssyca, Lina, Lucas, Manoel, Marcela, Tatiane, Thaisa e Rodolfo pela amizade e apoio de cada um durante mais uma etapa.

À minha orientadora, Thelma Sáfadi, pela atenção, confiança, conselhos e generosidade durante este período tão importante.

Aos docentes do programa de pós-graduação em Estatística e Experimentação Agropecuária da Universidade Federal de Lavras, pela dedicação e pelo conhecimento transmitido durante a pós-graduação.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## RESUMO

A análise de séries temporais a partir da classe de modelos autorregressivos e de médias móveis, ARMA(p, q), é amplamente adotada em estudos aplicados. Entretanto, séries de contagem necessitam de certa atenção pois podem apresentar características estilizadas, como alta dispersão e excesso de zeros, que devem ser levadas em conta pelo pesquisador. Ampliações da classe ARMA(p, q) para a modelagem de séries de contagem têm sido propostas na literatura estatística e talvez a mais difundida seja a classe de modelos autorregressivos e de médias móveis generalizados, GARMA(p, q), proposta no ano de 2003. Nesta tese, foram realizados três ensaios com o objetivo de estudar e estender alguns modelos da classe GARMA(p, q), a partir de diferentes distribuições e formas da estrutura de dependência temporal ao analisar contagens. No primeiro ensaio, realizou-se a análise utilizando as distribuições Poisson, a Binomial negativa e a Poisson inversa Gaussiana, assumindo que o parâmetro de dispersão da Binomial negativa é desconhecido e trazendo uma ampliação para o uso da Poisson inversa Gaussiana. No segundo ensaio, foram consideradas as versões zero-ajustadas das distribuições utilizadas no primeiro ensaio, estendendo a estrutura de dependência temporal de modo a possibilitar o uso de componentes sazonais e levar em conta fenômenos com muitos zeros. No terceiro ensaio, propôs-se o uso da distribuição Poisson zero-ajustada cujos parâmetros variam no tempo, permitindo a realização de previsões da série temporal e da probabilidade de contagens iguais a zero. Adotou-se a inferência Bayesiana dos modelos estudados nesta tese, sendo avaliados computacionalmente e utilizados em aplicações. Ademais, entende-se que as ampliações da classe GARMA(p, q) para versões que lidem com fenômenos sazonais e/ou com excesso de zeros e a adoção do algoritmo Monte Carlo Hamiltoniano para a amostragem da posteriori conjunta no terceiro ensaio são as principais contribuições deste trabalho. Os resultados dessa tese propiciam a análise de séries de contagem visando alternativas além da tradicional classe ARMA(p, q) e podem ser ampliados para outras distribuições de contagem e métodos de estimação.

**Palavras-chave:** ARMA(p, q). GARMA(p, q). Inferência Bayesiana.

## ABSTRACT

The analysis of time series based on the autoregressive and moving average class,  $ARMA(p, q)$ , is widely adopted in applied studies. However, count time series require some attention because they may have stylized characteristics, such as a high dispersion and an excess of zeros, which must be considered by the researcher. Extensions of this class for modeling count time series have been proposed in the statistical literature, and perhaps the most widespread is the generalized autoregressive moving average,  $GARMA(p, q)$ , proposed in 2003. In this thesis, we carried out three assays with the aim of studying and extending some models of the  $GARMA(p, q)$  class, from different distributions and forms of the temporal dependence structure when analyzing count time series. In the first assay, we used the Poisson, the negative Binomial, and the Poisson inverse Gaussian distributions, assuming that the negative Binomial dispersion parameter is unknown and bringing an extension to the Poisson inverse Gaussian distribution. In the second, the zero-adjusted versions of the distributions used in the first assay were considered, extending the temporal dependence structure in order to allow seasonal components and to take into account phenomena with many zeros. In the third, we propose the use of the zero-adjusted Poisson distribution whose parameters vary over time, allowing predictions of the time series and the probability of counts equal to zero. We adopted the Bayesian approach for inference, and the models were computationally evaluated and used in applications. Furthermore, we understand that the extensions of the  $GARMA(p, q)$  class for versions that deal with seasonal phenomena and/or with an excess of zeros, and the adoption of the Hamiltonian Monte Carlo algorithm for a sampling of the joint posterior in the third essay are the main contributions. The findings of this thesis allow the analysis of count time series aiming to look at alternatives beyond the  $ARMA(p, q)$  class, which can be extended to other count distributions and estimation methods.

**Keywords:**  $ARMA(p, q)$ . Bayesian Inference.  $GARMA(p, q)$ .

## SUMÁRIO

	<b>PRIMEIRA PARTE - UM PANORAMA GERAL</b> . . . . .	7
<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	9
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> . . . . .	11
<b>2.1</b>	<b>Modelos autorregressivos e de médias móveis</b> . . . . .	11
<b>2.2</b>	<b>Modelos autorregressivos e de médias móveis generalizados</b> . . . . .	12
<b>2.3</b>	<b>Abordagens para séries não Gaussianas</b> . . . . .	15
<b>2.4</b>	<b>Algumas distribuições de contagem</b> . . . . .	17
<b>2.4.1</b>	<b>Binomial negativa</b> . . . . .	18
<b>2.4.2</b>	<b>Poisson inversa Gaussiana</b> . . . . .	20
<b>2.4.3</b>	<b>Distribuições zero-ajustadas</b> . . . . .	22
<b>2.5</b>	<b>Aspectos inferenciais</b> . . . . .	24
<b>2.5.1</b>	<b>Distribuições a priori</b> . . . . .	25
<b>2.5.2</b>	<b>Métodos de Monte Carlo via cadeias de Markov</b> . . . . .	26
<b>2.5.2.1</b>	<b>Monte Carlo Hamiltoniano</b> . . . . .	28
<b>2.5.2.2</b>	<b>Monitoramento da convergência</b> . . . . .	31
<b>2.5.3</b>	<b>Estimação pontual e intervalar</b> . . . . .	34
<b>2.5.4</b>	<b>Comparação e adequação de modelos</b> . . . . .	34
<b>2.5.4.1</b>	<b>Crítério de informação da deviância</b> . . . . .	34
<b>2.5.4.2</b>	<b>Distribuição preditiva e a ordenada preditiva condicional</b> . . . . .	35
<b>3</b>	<b>CONCLUSÃO</b> . . . . .	36
	<b>REFERÊNCIAS</b> . . . . .	37
	<b>SEGUNDA PARTE - ARTIGOS</b> . . . . .	42
	<b>ARTIGO 1 - Analysis of count time series: a Bayesian GARMA(p, q) approach</b> . . . . .	43
	<b>ARTIGO 2 - Count time series with excess zeros: a Bayesian perspective using zero-adjusted distributions</b> . . . . .	65
	<b>ARTIGO 3 - Distribuição Poisson zero-ajustada com parâmetros variando no tempo para a análise de séries temporais de contagem</b> . . . . .	80
	<b>APÊNDICE A - Códigos</b> . . . . .	94



## **PRIMEIRA PARTE - UM PANORAMA GERAL**

## 1 INTRODUÇÃO

Ao revisitar a definição de séries temporais apresentada em Morettin e Toloi (2006) como conjuntos de observações ordenadas no tempo nos quais são estabelecidos modelos com objetivos diversos como o de previsão e o de análise de periodicidade, pode-se observar o constante interesse da literatura estatística em suprir e atender as particularidades de diferentes áreas aplicadas.

Na área financeira, por exemplo, modelos de volatilidade vêm sendo propostos e aprimorados para comportar as estruturas e as características estilizadas de séries temporais financeiras, que comumente possuem caudas pesadas e agrupamentos de volatilidade, possibilitando um ferramental para a tomada de decisão acerca da criação de portfólios e da alocação de recursos.

Em outra perspectiva, estudos como o de Barreto-Souza (2017) e o de Davis *et al.* (2021) apontaram que há um recente interesse na análise de séries temporais de contagem. De fato, isso pode ser constatado através das recentes e diversas publicações que envolvem o assunto, abrangendo aplicações em fenômenos atuariais, demográficos, econômicos e epidemiológicos (BRIËT; AMERASINGHE; VOUNATSOU, 2013; ZHA; LORD; ZOU, 2016; ASSIS, 2020).

Séries de contagem podem apresentar excessos de zeros, subdispersão e superdispersão, além da correlação que possa vir a existir ao longo do tempo e que comumente é não negativa (DAVIS *et al.*, 2021). Características como essas podem prejudicar o desempenho dos modelos tradicionais ao se analisar esses dados e, com isso, abordagens para a modelagem foram e têm sido desenvolvidas.

A justificativa para a realização desse estudo é a constante necessidade do aperfeiçoamento de modelos estatísticos para a análise de séries de contagem, de modo com que estes sejam mais condizentes com a realidade e forneçam alternativas para a análise de dados em estudos aplicados, tendo em vista as peculiaridades que podem existir.

O objetivo deste trabalho é estudar e estender alguns modelos para séries temporais de contagem sob a perspectiva da inferência Bayesiana, fazendo o uso de diferentes distribuições e estruturas de dependência temporal. Como objetivos específicos, busca-se adotar distribuições como a Poisson, a Binomial negativa e a Poisson inversa Gaussiana e as suas variações zero-

ajustadas, bem como permitir a inserção de componentes sazonais e parâmetros variando no tempo.

A estrutura dessa tese é composta por duas partes. Na primeira delas estão apresentados os tópicos que fundamentam a base teórica e na segunda parte tem-se os seguintes ensaios: (i) “*Analysis of count time series: a Bayesian GARMA( $p, q$ ) approach*”, aceito para publicação no *Austrian journal of statistics*; (ii) “*Count time series with excess zeros: a Bayesian perspective using zero-adjusted distributions*”, publicado no periódico *Semina: exact and technological sciences*; e (iii) "Distribuição Poisson zero-ajustada com parâmetros variando no tempo para a análise de séries temporais de contagem", que encontra-se em processo de submissão no *Journal of applied statistics*.

## 2 REFERENCIAL TEÓRICO

Nesta seção é apresentada a fundamentação teórica deste trabalho, abrangendo modelos para séries temporais, distribuições para a análise de dados de contagem e conceitos da inferência Bayesiana.

### 2.1 Modelos autorregressivos e de médias móveis

No estudo de séries temporais, modelos paramétricos ou não paramétricos são especificados com diversas finalidades como a realização de previsões e a análise de periodicidade (MORETTIN; TOLOI, 2006). Considerando modelos paramétricos, a metodologia de Box e Jenkins (1976) é vista como a mais difundida e utilizada em áreas aplicadas (SILVA, 2020).

Na metodologia de Box e Jenkins (1976), assume-se que o mecanismo gerador de uma série temporal é dado pela agregação linear de choques aleatórios, provenientes de variáveis aleatórias não correlacionadas com média zero e variância constante, usualmente denominadas como ruído branco ( $a_t$ ) e com distribuição normal. Esses ruídos também podem ser vistos como inovações no sistema.

De acordo com Morettin e Toloí (2006), essa ideia parte do processo linear geral em que uma série  $Z_t$ ,  $\{t = 1, \dots, n\}$ , é gerada por um filtro linear, isto é:

$$Z_t = \mu + \sum_{j=1}^{\infty} \psi_j a_{t-j}, \quad (2.1)$$

onde  $\mu$  denota o nível da série caso a série  $\sum_{j=1}^{\infty} \psi_j$  seja convergente. Uma segunda representação para a Equação (2.1) é dada por uma soma ponderada de desvios de  $Z_t$  em relação ao nível  $\mu$ , ou seja:

$$Z_t - \mu = \sum_{j=1}^{\infty} \phi_j (Z_{t-j} - \mu) + a_t.$$

Sob essa representação, pode-se definir um processo autorregressivo de ordem  $p$ , AR( $p$ ), se  $\phi_j = 0$  para  $j > p$ , isto é:

$$Z_t - \mu = \phi_1 (Z_{t-1} - \mu) + \dots + \phi_p (Z_{t-p} - \mu) + a_t.$$

Reescrevendo o processo AR(p) em termos do operador de traslado para o passado ( $B$ ) e adotando o polinômio autorregressivo de ordem  $p$ , dado por  $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ , tem-se que  $\phi(B)(Z_t - \mu) = a_t$ , sendo estacionário caso as raízes de  $\phi(B)$  estejam fora do círculo unitário e sempre será invertível. Um processo de médias móveis, MA(q), pode ser definido via Equação (2.1) caso  $\psi_j = 0$  para  $j > q$ , ou seja:

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q},$$

ou ainda  $Z_t - \mu = \theta(B)a_t$ , em que  $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ , sendo invertível se as raízes de  $\theta(B)$  estiverem fora do círculo unitário e sempre será estacionário. O processo MA(q) pode ser visto como uma saída  $Z_t - \mu$  de um filtro linear com função de transferência  $\theta(B)$  cuja entrada é um ruído branco (BOX; JENKINS, 1976).

Os modelos AR(p) e MA(q) podem necessitar de um elevado número de parâmetros para descrever a dinâmica da série temporal (TSAY, 2013). Uma solução mais parcimoniosa é a combinação destes dois processos, resultando no processo autorregressivo e de médias móveis, ARMA(p, q) (MORETTIN; TOLOI, 2006), que pode ser denotado da seguinte forma:

$$\phi(B)(Z_t - \mu) = \theta(B)a_t.$$

Para o processo ARMA(p, q), as condições de estacionariedade e de invertibilidade são atendidas se as raízes de  $\phi(B)$  e  $\theta(B)$  estiverem fora do círculo unitário, respectivamente. Ademais, formas de identificação das ordens  $p$  e  $q$ , métodos de inferência e extensões para modelos sazonais estão em Morettin e Toloí (2006), Tsay (2013) e Morettin (2017).

Segundo Morettin (2017), os modelos lineares AR(p), MA(q) e ARMA(p, q) são apropriados para a análise de séries Gaussianas; mas são muito utilizados e possibilitam realizar previsões satisfatórias e eficientes (SILVA, 2020). Todavia, uma alternativa para a análise de séries não Gaussianas, vista como uma extensão da classe ARMA(p, q), tem sido desenvolvida na literatura estatística e está apresentada na subseção 2.2.

## 2.2 Modelos autorregressivos e de médias móveis generalizados

A classe de modelos autorregressivos e de médias móveis generalizados (GARMA) foi proposta por Benjamin, Rigby e Stasinopoulos (2003), assumindo que uma série temporal  $Y$  é modelada por uma distribuição pertencente à família exponencial e o preditor linear é composto

por um conjunto de variáveis explicativas e uma estrutura de dependência na forma ARMA(p, q).

Denotando  $Y = \{y_1, \dots, y_n\}$  como o conjunto de observações da série temporal;  $\mathcal{F}_{t-1}$  como o conjunto de informações passadas até o instante  $t - 1$ , para  $t = \{1, \dots, n\}$ , dado por  $\mathcal{F}_{t-1} = \{y_1, y_2, \dots, y_{t-1}, \mu_1, \mu_2, \dots, \mu_{t-1}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{t-1}\}$ , sendo  $\mathbf{x}_t$  o vetor de variáveis explicativas de tamanho  $r$ , a densidade condicional de cada observação  $y_t$  é expressa como mostrado na Equação (2.2):

$$p(y_t | \mathcal{F}_{t-1}) = \exp \left[ \frac{y_t \vartheta_t - b(\vartheta_t)}{\varphi} + d(y_t, \varphi) \right], \quad (2.2)$$

em que  $\vartheta_t$  é o parâmetro canônico,  $\varphi$  é o parâmetro de escala e  $b(\cdot)$  e  $d(\cdot)$  são funções similares às dos modelos lineares generalizados de Nelder e Wedderburn (1972).

Como a distribuição de  $y_t$  é condicionada ao conjunto  $\mathcal{F}_{t-1}$ , tem-se que  $\mu_t = b'(\vartheta_t) = E(y_t | \mathcal{F}_{t-1})$  e  $V(y_t | \mathcal{F}_{t-1}) = \varphi b''(\vartheta_t)$ . A média condicional é relacionada ao preditor linear ( $\eta_t$ ) através de uma função de ligação  $g(\cdot)$  que é monótona e diferenciável de segunda ordem (BENJAMIN; RIGBY; STASINOPOULOS, 2003). Uma forma geral para  $\eta_t$  é mostrada na Equação (2.3):

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

em que  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_r)^\top$  é o vetor de parâmetros associado as  $r$  variáveis explicativas e  $\tau_t$  é o termo acrescido de forma aditiva à  $\eta_t$ , contendo a estrutura de dependência.

O componente  $\tau_t$  foi definido em termos autorregressivos e de médias móveis como apresentado na Equação (2.4):

$$\tau_t = \sum_{j=1}^p \phi_j A(y_{t-j}, \mathbf{x}_{t-j}, \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j M(y_{t-j}, \mu_{t-j}), \quad (2.4)$$

em que  $A$  e  $M$  são funções que representam as estruturas AR(p) e MA(q), respectivamente, com vetores paramétricos  $\boldsymbol{\Phi} = (\phi_1, \phi_2, \dots, \phi_p)^\top$  e  $\boldsymbol{\Theta} = (\theta_1, \theta_2, \dots, \theta_q)^\top$ , ou ainda:

$$\tau_t = \sum_{j=1}^p \phi_j [g(y_{t-j}) - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j [g(y_{t-j}) - \eta_{t-j}]. \quad (2.5)$$

Para a representação (2.5) pode ser necessário substituir  $y_{t-j}$  por  $y_{t-j}^*$  como forma de garantir a existência da função de ligação. O que se tem visto em alguns estudos é a utilização de  $y_{t-j}^* = \max(y_{t-j}, c)$ ,  $c \in (0, 1)$ , para o caso em que a função de ligação é o logaritmo natural ou em que a função  $g(\cdot)$  tem domínio em  $\mathbb{R}^+$ .

O procedimento de estimação de parâmetros dos modelos GARMA( $p, q$ ) foi descrito em Benjamin, Rigby e Stasinopoulos (2003) utilizando a função de verossimilhança condicionada às  $m = \max(p, q)$  primeiras observações, com maximização obtida numericamente. Em alternativa, Andrade, Andrade e Ehlers (2015) e Franco, Migon e Marcos (2015) adotaram a análise Bayesiana com a utilização de algoritmos para amostragem da posteriori conjunta.

Estudos utilizam os resíduos quantílicos randomizados normalizados para verificar a adequação dos modelos, que são dados por  $r_t = \Phi^{-1}(u_t)$ ,  $t = \{1, \dots, n\}$ , onde  $\Phi^{-1}(\cdot)$  representa a inversa acumulada da normal padrão e  $u_t \sim U[F(y_t - 1, \mathcal{F}_{t-1}); F(y_t, \mathcal{F}_{t-1})]$ , sendo  $F(\cdot)$  a distribuição condicional acumulada. Ou seja,  $r_t$  é construído invertendo a distribuição ajustada e encontrando o quantil correspondente na normal padrão. Com isso, sob condições de que o modelo assumido está correto,  $r_t \rightarrow N(0, 1)$  (DUNN; SMYTH, 1996).

Stasinopoulos *et al.* (2017) apontam que quando o processo de aleatorização dos resíduos é utilizado, vários conjuntos de resíduos, ou sua média, devem ser analisados previamente antes de se tomar a decisão sobre a adequação do modelo, isso pois o valor do resíduo quantílico se modifica para cada realização da variável  $U[F(y_t - 1, \mathcal{F}_{t-1}); F(y_t, \mathcal{F}_{t-1})]$ .

Os diagnósticos de  $r_t$  são realizados por meio de testes de normalidade, análise do histograma e testes para as autocorrelações por meio das estatísticas de Box e Pierce (1970) ou Ljung e Box (1978). Como visto em Morettin (2017), a estatística de Ljung e Box (1978) é dada por:

$$Q(K) = n(n+2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)},$$

sendo  $\hat{r}_j$  as autocorrelações dos resíduos estimados até as  $K$  primeiras defasagens. A hipótese nula de ruído branco é rejeitada para valores de  $Q(K) > \chi_{K-p-q}^2$ .

Condições de estacionariedade da classe foram avaliadas em cenários em que  $p(y_t | \mathcal{F}_{t-1})$  pertence à família exponencial e  $g(\cdot)$  é a identidade. Para isso, é necessário garantir a invertibilidade do polinômio característico, ou seja, certificar que as raízes do polinômio estão fora do círculo unitário (MELO; ALENCAR, 2020). Em outras situações, definir tais condições pode ser uma tarefa complicada (BENJAMIN; RIGBY; STASINOPOULOS, 2003).

Posteriormente a criação da classe GARMA( $p, q$ ), outros modelos e ampliações foram desenvolvidos e incorporados à literatura para a análise de séries temporais não Gaussianas. Parte desses estudos adotaram diferentes distribuições de probabilidade e/ou estruturas de dependência temporal e estão brevemente apresentados na subseção 2.3.

### 2.3 Abordagens para séries não Gaussianas

Shephard (1995) propôs a classe *Generalized Linear Autoregressive Moving Average* (GLARMA), que foi estudada e ampliada por Davis, Dunsmuir e Streett (2003), mantendo a condição de que a distribuição condicional a  $\mathcal{F}_{t-1}$  pertença à família exponencial na forma:  $p(y_t | W_t) = \exp\{y_t W_t - a_t b(W_t) + c_t\}$ , sendo  $a_t$  e  $c_t$  constantes e  $W_t$  uma variável que resume  $\mathcal{F}_{t-1}$ . Desta forma,  $W_t = \mathbf{x}_t^\top \boldsymbol{\beta} + Z_t$ , em que  $Z_t$  induz a dependência serial nos estados (DUNSMUIR, 2016).

Rocha e Cribari-Neto (2008) apresentaram o modelo *Beta Autoregressive Moving Average* ( $\beta$ ARMA) cuja estrutura de dependência temporal é similar à do GARMA(p, q), para a modelagem de séries que assumem valor no intervalo unitário. O  $\beta$ ARMA foi ampliado no estudo de Bayer, Cintra e Cribari-Neto (2018) com a adição de componentes sazonais, sendo denominado como  $\beta$ SARMA.

Briët, Amerasinghe e Vounatsou (2013) propuseram o *Generalized Seasonal Autoregressive Integrated Moving Average* (GSARIMA) considerando a distribuição Binomial negativa para a análise dos casos de Malária reportados no *Sri Lanka*. Os autores adotaram a abordagem Bayesiana e verificaram que o modelo obteve melhor desempenho quando comparado ao Gaussiano.

Franco, Migon e Marcos (2015) compararam os modelos com dependência do tipo *observation driven* e *parameter driven*<sup>1</sup>, assumindo a distribuição Poisson para a resposta. Verificou-se melhor desempenho do modelo com estrutura do tipo *parameter driven* em regiões do espaço paramétrico próximas a condição de não estacionariedade, mas apresentando alguns problemas de autocorrelação nos resíduos.

Andrade, Andrade e Ehlers (2015) utilizaram a inferência Bayesiana para o estudo de modelos GARMA(p, q) com distribuições Binomial, Binomial negativa e Poisson. Os modelos foram construídos com prioris não informativas e a amostragem foi realizada por métodos de Monte Carlo via cadeias de Markov. Os autores apontaram que tal inferência, embora ainda pouco utilizada nesses modelos, apresentou ganhos em termos de estimação e de amplitude de intervalos de credibilidade.

Ehlers (2019) propôs o modelo *Conway-Maxwell Poisson* para a análise de séries temporais de contagem, inserindo a estrutura de dependência da classe GARMA(p, q) e considerando que o parâmetro de dispersão também varia no tempo e depende de defasagens da série. O

<sup>1</sup> Definições dessas estruturas encontram-se em Davis *et al.* (2021).



autor realizou uma análise Bayesiana com o uso de prioris não informativas e exemplificou a utilização do modelo em dados de poliomelite.

O GARMA modificado (GAR-M(p, q)), foi apresentado por Albarracin, Alencar e Ho (2019) com o objetivo de reduzir a multicolinearidade propiciada pela estrutura do preditor dos GARMA(p, q), que pode resultar em falhas de convergência de algoritmos de estimação dos parâmetros. Observou-se uma redução da multicolinearidade e, além disso, os autores apontaram que o GARMA(p, q) tradicional com estrutura AR(p) ou MA(q) é suficiente para modelar a correlação serial.

Silva (2020) apresentou a classe *Generalized Autoregressive Neural Network*, em que  $\tau_t$  foi substituído por uma rede neural com a finalidade de comportar a correlação serial. Defasagens da série e do componente sistemático foram considerados como variáveis de entrada da rede. Também foram apresentadas aplicações para a análise de casos de poliomelite em modelos com resposta Binomial negativa e Poisson.

Na temática de séries temporais resultantes de contagens, alguns autores direcionaram-se a análise de fenômenos com muitos zeros. Exemplos destes são os estudos de Yang, Zamba e Cavanaugh (2013), Alqawba, Diawara e Chaganty (2019), Assis (2020), Ghahramani e White (2020), Sathish, Mukhopadhyay e Tiwari (2021) e Tawiah, Iddrisu e Asosega (2021).

Yang, Zamba e Cavanaugh (2013) utilizaram a regressão Poisson inflada de zeros, também denominada *Zero-Inflated Poisson (ZIP)*, inserindo defasagens da série para tratar a correlação serial. Os autores conduziram um estudo de simulação, realizaram uma aplicação em casos de sífilis e argumentaram que a metodologia pode ser ampliada para outras distribuições, como a Binomial negativa inflada de zeros (ZINB).

Distribuições como a ZIP e a ZINB foram empregues por Alqawba, Diawara e Chaganty (2019), sendo suposto que os erros das regressões seguiam um processo ARMA(p, q) estacionário. Um processo de simulação foi realizado em cenários com séries temporais de tamanho 100, 200 e 500 e a metodologia foi exemplificada ao se analisar os dados de acidentes de trabalho ocorridos em um hospital.

Assis (2020) abordou modelos Bayesianos do tipo zero-modificados com estrutura de dependência ARMA(p, q). No estudo, foram realizadas simulações e aplicações em dados de rubéola e de sífilis. Os modelos Poisson e COM-Poisson zero-modificados mostraram-se interessantes para a análise e, segundo os autores, podem ser estendidos para o contexto de séries sazonais e/ou multivariadas.

Sathish, Mukhopadhyay e Tiwari (2021) consideraram modelos inflados de zero com estrutura de dependência ARMA(p, q) e com a introdução de variáveis explicativas no preditor linear para o controle de tendência e de variações sazonais (retratadas por funções seno e cosseno). Os autores realizaram um estudo de simulação e analisaram as séries temporais de casos semanais de dengue e de sífilis.

Os modelos ZIP e ZINB foram utilizados por Tawiah, Iddrisu e Asosega (2021) para analisar a série temporal de mortalidade em decorrência da COVID-19 em Gana, que apresentava muitos zeros durante o início da pandemia. Foram realizadas previsões e comparações do número de óbitos pela causa, permitindo também a análise de tendências para os meses de novembro e dezembro de 2020.

A partir da exposição dos estudos que analisaram séries não Gaussianas, pode-se observar que, ao lidar com contagens, algumas distribuições difundidas na literatura como a Binomial negativa e a Poisson foram utilizadas. Sendo assim, na subseção 2.4 estão apresentadas algumas das diversas distribuições que permitem modelar contagens.

## 2.4 Algumas distribuições de contagem

Dados de contagem surgem naturalmente em estudos aplicados e observa-se um recente interesse na modelagem destes (BARRETO-SOUZA, 2017). Na maioria das situações práticas, os estudos recorrem ao uso da Poisson,  $POI(\mu)$ , fato que pode ser justificado pela facilidade de implementação e de tratabilidade algébrica (GONÇALVES; BARRETO-SOUZA, 2020), cuja função de probabilidade é dada por:

$$p(Y = y | \mu) = \frac{\exp(-\mu)\mu^y}{y!},$$

definida no espaço paramétrico  $\Omega = \{\mu, \mu > 0\}$  e a característica de que  $E(Y) = V(Y) = \mu$ . Ao abordar dados temporalmente correlacionados, deriva-se o submodelo GARMA Poisson inserindo o preditor mostrado na Equação (2.3) em  $\mu$  (BENJAMIN; RIGBY; STASINOPOULOS, 2003).

A  $POI(\mu)$  pode não ser adequada para dados com alta dispersão, que é um fenômeno constantemente observado em contagens e pode ser resultante de excesso de zeros (BARRETO-SOUZA, 2017), da presença de observações atípicas e da heterogeneidade da população (PAYNE

*et al.*, 2017). Para lidar com tal situação e permitir analisar dados mais dispersos, distribuições de mistura podem ser utilizadas.

Uma variável aleatória  $Y$  segue uma mistura  $\text{POI}(\mu)$  se satisfizer a representação estocástica de  $Y | Z = z \sim \text{POI}(\mu z)$ , em que  $Z$  é uma variável aleatória não negativa. A distribuição de  $Y$  é obtida por meio da integração (2.6):

$$p(Y = y | \mu) = \int_0^{\infty} \frac{\exp(-\mu z)(\mu z)^y}{y!} dG_{\vec{\phi}}(z), \quad (2.6)$$

sendo  $G_{\vec{\phi}}(\cdot)$  a função de distribuição e  $\vec{\phi}$  o vetor paramétrico de  $Z$ . A partir da função geradora de momentos pode-se obter que a  $E(Y) = \mu E(Z)$  e a  $V(Y) = \mu[E(Z) + \mu V(Z)]$  (BARRETO-SOUZA, 2017). Em 2.4.1 e 2.4.2 estão apresentadas duas distribuições de mistura que possibilitam a análise de dados mais dispersos.

#### 2.4.1 Binomial negativa

Segundo Hilbe (2014), a Binomial negativa tradicional pode ser derivada a partir da mistura das distribuições  $\text{POI}(\mu)$  e Gama,  $\text{GA}(u, v)$ , mas essa não é a única forma encontrada na literatura. Uma origem alternativa é dada a partir da função de probabilidades da binomial. Essas variações podem resultar em diferentes parametrizações da distribuição, sendo necessário e importante definir qual está sendo utilizada (OESSELMANN, 2016).

Na derivação via mistura, assume-se na Equação (2.6) que a variável  $Z \sim \text{GA}(u = 1, v)$ , cuja função densidade de probabilidade é dada por:

$$p(z | u, v) = \frac{1}{\Gamma(v)} \left(\frac{zv}{u}\right)^v \exp\left(-\frac{vz}{u}\right) \left(\frac{1}{z}\right),$$

definida em  $\Omega = \{u, v | u, v > 0\}$ , com  $E(Z) = u = 1$  e  $V(Z) = \frac{1}{v}$ . A distribuição de  $Y$ , obtida pela integração em respeito à  $z$ , é denominada Binomial negativa,  $\text{BN}(\mu, v)$ , apresentada na Equação (2.7):

$$p(Y = y | \mu, v) = \frac{\Gamma(v + y)}{\Gamma(v)\Gamma(y + 1)} \left(\frac{\mu}{\mu + v}\right)^y \left(\frac{v}{\mu + v}\right)^v, \quad (2.7)$$

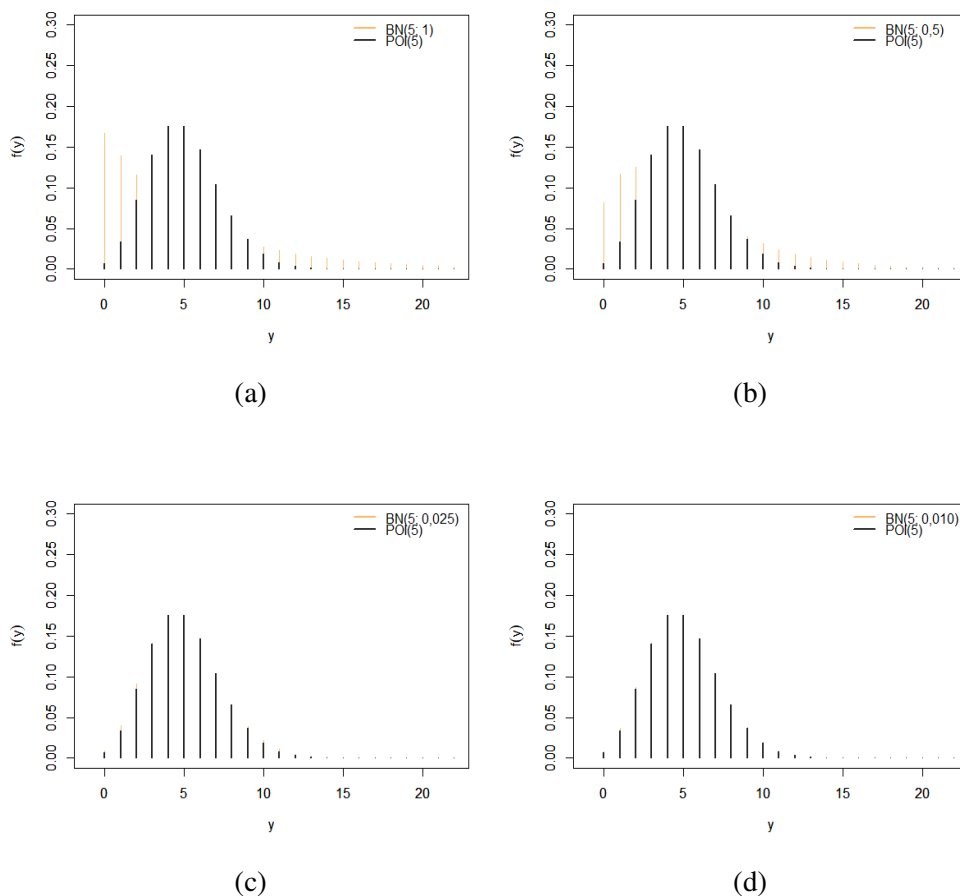
para  $y \in \mathbb{N}$ , podendo ser fatorada na forma da família exponencial assumindo que  $v$  é conhecido (OESSELMANN, 2016). Passos algébricos da integração mostrada na Equação (2.6) estão em Hilbe (2014).

Hilbe (2014) propõe inverter o parâmetro de escala de  $Z$  gerando  $\sigma = v^{-1}$ , visto como um parâmetro inversamente relacionado com a superdispersão dos dados e sendo dificilmente maior que 4. Reescrevendo a Equação (2.7) em termos de  $\sigma$ , tem-se que:

$$p(Y = y | \mu, \sigma) = \frac{\Gamma(\frac{1}{\sigma} + y)}{\Gamma(\frac{1}{\sigma}) \Gamma(y + 1)} \left( \frac{\sigma \mu}{\sigma \mu + 1} \right)^y \left( \frac{1}{\sigma \mu + 1} \right)^{\frac{1}{\sigma}},$$

que corresponde a parametrização adotada por Stasinopoulos e Rigby (2020) e simbolizada por  $BN(\mu, \sigma)$ . Sob essa forma a  $E(Y) = \mu$  e a  $V(Y) = \mu + \mu^2 \sigma$ , fazendo com que a  $POI(\mu)$  seja a distribuição limite a medida que  $\sigma \rightarrow 0$ . Na Figura 2.1 estão apresentados dois cenários de comparação entre as distribuições  $POI(\mu)$  e  $BN(\mu, \sigma)$ .

Figura 2.1 – Comparativo entre a  $BN(\mu, \sigma)$  e a  $POI(\mu)$  em diferentes cenários de  $\sigma$ , considerando  $\mu = 5$ .



Fonte: Do autor (2023).

Ao assumir a distribuição Binomial negativa na classe  $GARMA(p, q)$ , Benjamin, Rigby e Stasinopoulos (2003) adotaram a parametrização mostrada na Equação (2.7) inserindo o pre-

ditor (2.3) com função de ligação logarítmica em  $\mu$ . A mesma parametrização foi utilizada por Andrade, Andrade e Ehlers (2015), reescrevendo a distribuição como um membro da família exponencial em que  $k = v^{-1}$  é conhecido.

#### 2.4.2 Poisson inversa Gaussiana

A Poisson inversa Gaussiana,  $PIG(\mu, \sigma)$ , é uma mistura entre as distribuições  $POI(\mu)$  e a inversa Gaussiana,  $IG(u, \sigma)$ , vista como flexível e útil na modelagem de dados de contagem com caudas pesadas e curtose elevada (ZHA; LORD; ZOU, 2016; HELLER; COUTURIER; HERITIER, 2018). É um caso especial da distribuição de *Sichel*, sendo uma alternativa para fenômenos com alta dispersão, assim como a  $BN(\mu, \sigma)$  (OUMA; MWALILI; KIBERIA, 2016).

Abordando a notação de Zha, Lord e Zou (2016), considera-se que a função densidade de probabilidade de  $Z$  é dada por:

$$p(z | u, \sigma) = \frac{1}{\sqrt{2\pi\sigma z^3}} \exp \left[ \frac{-(z-u)^2}{2u^2\sigma z} \right],$$

para  $\Omega = \{u, \sigma | u, \sigma > 0\}$ , em que a  $E(Z) = u = 1$  e a  $V(Z) = \sigma u^3$ . A distribuição de  $Y$ , obtida pela integração da Equação (2.6) em respeito à  $z$ , é denominada  $PIG(\mu, \sigma)$  e está apresentada na Equação (2.8):

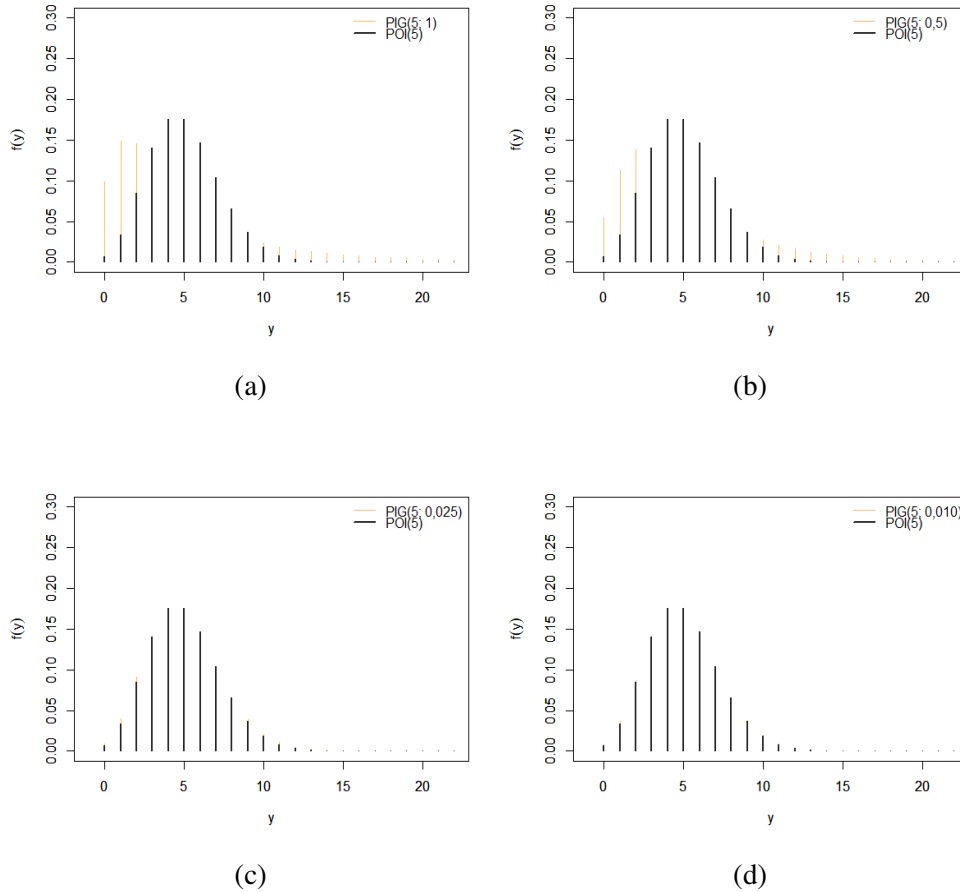
$$p(Y = y | \mu, \sigma) = \left( \frac{2\alpha}{\pi} \right)^{\frac{1}{2}} \mu^y e^{\frac{1}{\sigma}} \left( K_{y-\frac{1}{2}}(\alpha) \right) \frac{1}{(\alpha\sigma)^y y!}, \quad (2.8)$$

sendo  $\alpha^2 = \frac{1}{\sigma^2} + \frac{2\mu}{\sigma}$  e  $K_\lambda(t) = \frac{1}{2} \int_0^\infty x^{\lambda-1} \exp(-\frac{t}{2}(x+x^{-1})) dx$ , em que  $K(\cdot)$  é a função de *Bessel* modificada do tipo II <sup>2</sup>. Definida em  $\Omega = \{\mu, \sigma | \mu, \sigma > 0\}$  para  $y \in \mathbb{N}$  (RIBGY *et al.*, 2019). A partir da função geradora de momentos, verifica-se que a  $E(Y) = \mu$  e a  $V(Y) = \mu(1 + \mu\sigma)$  (BARRETO-SOUZA, 2017).

O parâmetro  $\sigma$  atua inflando a variância de  $Y$ , sendo a  $POI(\mu)$  a distribuição limite conforme  $\sigma \rightarrow 0$  (RIBGY *et al.*, 2019). Verifica-se também que conforme  $\sigma \rightarrow 0$ , o coeficiente de assimetria da  $PIG(\mu, \sigma)$  é equivalente ao da  $POI(\mu)$ . Na Figura 2.2 estão apresentados cenários para a comparação dessas duas distribuições a medida em que  $\sigma$  tende a zero, considerando valores fixos de  $\mu = 5$ .

<sup>2</sup> Nos artigos 1 e 2, leia função de *Bessel* modificada do tipo II.

Figura 2.2 – Comparativo entre a  $PIG(\mu, \sigma)$  e a  $POI(\mu)$  em diferentes cenários de  $\sigma$ , considerando  $\mu = 5$ .



Fonte: Do autor (2023).

Ao abordar dados temporalmente correlacionados, Barreto-Souza (2017) propôs a utilização da  $PIG(\mu, \sigma)$  a partir do processo *Integer Valued Autoregressive* (INAR). Conforme Lima (2013), o  $INAR(p)$  é um processo de contagem para inteiros, dado por:

$$Y_t = \sum_{j=1}^p \alpha_j \circ Y_{t-j} + I_t, \quad \alpha_j \in (0, 1),$$

sendo baseado na combinação linear de operadores binomial *thinning*, em que  $I_t$  é uma variável aleatória discreta não negativa, representando a inovação, e  $Y_t$  é uma série temporal.

No  $INAR(p)$ , cada ordem de defasagem ( $j$ ) possui uma probabilidade  $\alpha_j$  a ser estimada e o procedimento de estimação é mais complexo quando comparado com os modelos  $AR(p)$ , dado que a distribuição de  $Y_t$  é uma convolução de  $I_t$  e do somatório dos binomiais com parâmetros  $Y_{t-j}$  e  $\alpha_j$  (LIMA, 2013). Maiores detalhes sobre o  $INAR(p)$  e das metodologias para a estimação

de parâmetros, como o de máxima verossimilhança, estão em Silva, Pereira e Silva (2009) e Lima (2013).

Como dito, um tópico com certa relevância em dados de contagem relaciona-se com a análise de fenômenos com muitos zeros. Neste caso, as distribuições comumente adotadas para a modelagem como a Binomial negativa e a Poisson podem não comportar essa situação (FENG, 2021). Esse fato eleva o interesse pela utilização de variações/ampliações destas distribuições, como as zero-ajustadas, que também podem ser denominadas como *hurdle* ou zero-alteradas.

### 2.4.3 Distribuições zero-ajustadas

Em distribuições zero-ajustadas, um modelo binomial governa o evento binário quando a resposta assume o valor zero ou positivo. Se o evento é positivo, a barreira é cruzada e o processo é modelado por uma distribuição condicional truncada em zero (MULLAHY, 1986). Conforme Feng (2021), a estrutura geral dessas distribuições é dada da seguinte forma:

$$P(Y = y | \boldsymbol{\theta}) = \begin{cases} v, & y = 0 \\ (1 - v) \frac{p(Y=y|\boldsymbol{\theta})}{1 - p(Y=0|\boldsymbol{\theta})}, & y > 0 \end{cases}, \quad (2.9)$$

em que  $v$  é a probabilidade de que  $Y$  seja zero,  $p(y | \boldsymbol{\theta})$  é a função de probabilidade de uma distribuição de contagem com vetor paramétrico  $\boldsymbol{\theta}$ .

Na Equação (2.9), a primeira estrutura modela a ocorrência de zeros com probabilidade  $v$  e a segunda é responsável pela modelagem de quando o zero não ocorre, cuja probabilidade é  $1 - v$  (FENG, 2021). Observe que os zeros não são diferenciados como ocorre nos modelos inflados de zero. Na verdade, os eventos que resultam em zero são vistos como provenientes de uma fonte estrutural (ZUUR *et al.*, 2009).

Há inúmeras possibilidades de se especificar a distribuição condicional dos valores positivos e do evento binário (MULLAHY, 1986). Uma delas é considerar que o evento binário é modelado por uma distribuição Binomial e os valores positivos seguem uma Poisson, resultando na distribuição Poisson zero-ajustada, do inglês *zero-adjusted Poisson* (ZAP) (HASHIM; HASHIM; SHIKER, 2021).

A seguir, estão apresentadas as versões zero-ajustadas das distribuições Binomial negativa, Poisson e Poisson inversa Gaussiana, adotando a parametrização de Rigby *et al.* (2019).

Para a distribuição Poisson zero-ajustada, a função de probabilidade é dada por:

$$p(Y = y | \mu, \nu) = \nu I_{(y=0)} + \left[ \frac{1 - \nu}{1 - e^{-\mu}} \right] e^{-\mu} \mu^y \frac{1}{y!} I_{(y>0)}, \quad (2.10)$$

para  $y \in \mathbb{N}$  e  $\Omega = \{\mu, \nu | \mu > 0, 0 < \nu < 1\}$ . Na Equação (2.10),  $I_{(\cdot)}$  representa a função indicadora,  $\nu$  modela a probabilidade exata de que a contagem seja zero e  $\mu$  é a esperança condicional da variável na situação em que  $y > 0$ .

Para a distribuição Binomial negativa zero-ajustada, ZANBI, a função de probabilidade é expressa da seguinte forma:

$$p(Y = y | \mu, \sigma, \nu) = \nu I_{(y=0)} + \frac{(1 - \nu)\mu}{[1 - (1 + \mu\sigma)^{-\frac{1}{\sigma}}]} \times \frac{\Gamma(y + \frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(y+1)} \left( \frac{\sigma\mu}{1 + \sigma\mu} \right)^y \left( \frac{1}{1 + \sigma\mu} \right)^{\frac{1}{\sigma}} I_{(y>0)},$$

definida em  $\Omega = \{\mu, \sigma, \nu | \mu, \sigma > 0, 0 < \nu < 1\}$  com  $y \in \mathbb{N}$ . O parâmetro  $\nu$  representa a probabilidade exata de que a contagem seja zero e os parâmetros  $\mu$  e  $\sigma$  modelam, respectivamente, a média e dispersão do componente positivo.

Para o caso da Poisson inversa Gaussiana zero-ajustada, ZAPIG, tem-se a seguinte função de probabilidade:

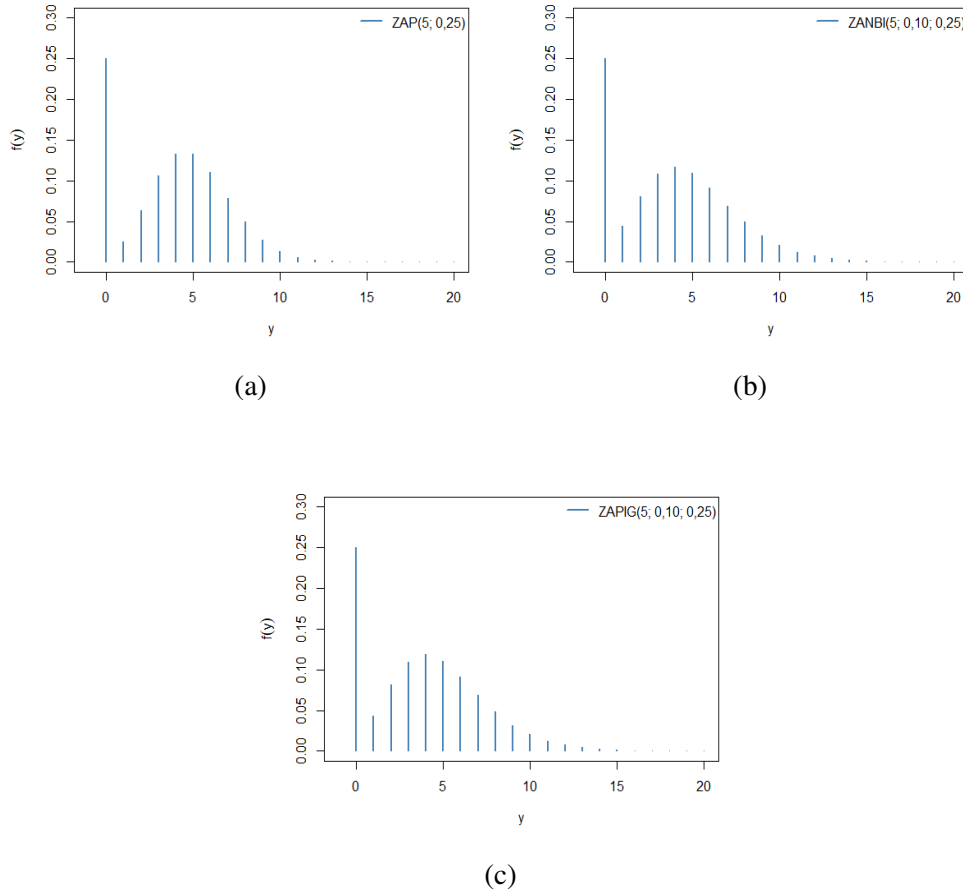
$$p(Y = y | \mu, \sigma, \nu) = \nu I_{(y=0)} + \frac{1 - \nu}{1 - e^{\frac{1}{\sigma} - \alpha}} \left( \frac{2\alpha}{\pi} \right)^{\frac{1}{2}} \mu^y e^{\frac{1}{\sigma}} \left( K_{y-\frac{1}{2}}(\alpha) \right) \frac{1}{(\alpha\sigma)^y y!} I_{(y>0)}, \quad (2.11)$$

sendo  $\Omega = \{\mu, \sigma, \nu | \mu, \sigma > 0, 0 < \nu < 1\}$  para  $y \in \mathbb{N}$ . Em (2.11)  $\alpha^2 = \frac{1}{\sigma^2} + \frac{2\mu}{\sigma}$  e  $K_\lambda(T) = \frac{1}{2} \int_0^\infty x^{\lambda-1} \exp\left\{-\frac{T}{2}(x+x^{-1})\right\} dx$ , sendo  $K(\cdot)$  a função de *Bessel* modificada do tipo II. Assim como na ZANBI,  $\nu$  expressa a probabilidade exata de que a contagem seja zero e  $\mu$  e  $\sigma$  são, respectivamente, a média e a dispersão do componente positivo.

Um comportamento das distribuições ZAP( $\mu, \nu$ ), ZANBI( $\mu, \sigma, \nu$ ) e ZAPIG( $\mu, \sigma, \nu$ ) está apresentado na Figura 2.3 considerando a situação de 25% de contagens iguais a zero. Além disso, outras distribuições do tipo zero-ajustadas e as respectivas expressões de esperança e variância das distribuições tratadas neste texto estão disponíveis em Rigby *et al.* (2019) e Zuo *et al.* (2021).



Figura 2.3 – Comparativo entre a  $ZAP(\mu, \nu)$ , a  $ZANBI(\mu, \sigma, \nu)$  e a  $ZAPIG(\mu, \sigma, \nu)$ , em (a), (b) e (c), respectivamente, considerando  $\mu = 5$ ,  $\nu = 0,25$  e  $\sigma = 0,10$ .



Fonte: Do autor (2023).

## 2.5 Aspectos inferenciais

No estudo de Benjamin, Rigby e Stasinopoulos (2003), a estimação dos parâmetros dos modelos  $GARMA(p, q)$  foi realizada por máxima verossimilhança utilizando o método de mínimos quadrados ponderados. Andrade, Andrade e Ehlers (2015) apresentaram uma abordagem através da inferência Bayesiana, apontando contribuições inferenciais, que está apresentada nesta seção.

Conforme Broemeling (2019), a inferência Bayesiana em séries temporais consiste em estimar, testar hipóteses acerca de parâmetros populacionais e fazer previsões através de distribuições preditivas. A fundamentação da teoria Bayesiana é dada no Teorema de *Bayes*, baseado na regra de probabilidades condicionais.

Para descrever o Teorema, Box e Tiao (1992) supõem um conjunto de observações  $\mathbf{y} = (y_1, \dots, y_n)^\top$  em que a distribuição de probabilidade  $p(\mathbf{y} | \boldsymbol{\theta})$  depende do vetor de parâmetros

$\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^\top$ , supondo também que  $\boldsymbol{\theta}$  tenha de uma distribuição de probabilidade denominada  $p(\boldsymbol{\theta})$ . Então, dada a relação:

$$p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta}) = p(\mathbf{y}, \boldsymbol{\theta}) = p(\boldsymbol{\theta} | \mathbf{y})p(\mathbf{y}),$$

a distribuição condicional de  $\boldsymbol{\theta}$  após observar  $\mathbf{y}$ , é dada por:

$$p(\boldsymbol{\theta} | \mathbf{y}) = \frac{p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})},$$

em que  $p(\mathbf{y})$  é escrita como  $\int p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$  caso  $\boldsymbol{\theta}$  seja contínuo e  $\sum p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})$  no caso discreto, sendo entendida como uma constante normalizadora. Dada a amostra  $\mathbf{y}$ , pode-se considerar  $p(\mathbf{y} | \boldsymbol{\theta})$  como uma função de  $\boldsymbol{\theta}$ , gerando a função de verossimilhança ( $L(\boldsymbol{\theta} | \mathbf{y})$ ). Com isso, a  $p(\boldsymbol{\theta} | \mathbf{y})$  pode ser escrita da seguinte forma:

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto L(\boldsymbol{\theta} | \mathbf{y})p(\boldsymbol{\theta}),$$

ou seja, uma relação de proporcionalidade entre a verossimilhança e a distribuição de  $\boldsymbol{\theta}$  (BOX; TIAO, 1992), resultando na distribuição a posteriori ( $p(\boldsymbol{\theta} | \mathbf{y})$ ) em que todas as inferências são realizadas (BROEMELING, 2019)<sup>3</sup>. A análise é uma forma de atualização do conhecimento prévio acerca de  $\boldsymbol{\theta}$  após a observação dos dados, cujo resultado da atualização está contido na posteriori.

O conhecimento prévio resumido probabilisticamente em  $\boldsymbol{\theta}$  é um ponto que recebe certo nível de críticas na análise Bayesiana, podendo ser especificado de diversas formas como na utilização de prioris conjugadas (WANG; YUE; FARAWAY, 2018).

### 2.5.1 Distribuições a priori

O processo de especificação de uma distribuição a priori é de grande importância na análise Bayesiana e talvez este seja o principal ponto de críticas do método. Um incentivo para a escolha de uma priori é o conhecimento sobre  $\boldsymbol{\theta}$ , que pode ser baseado na informação do pesquisador ou em resultados de experimentos passados (WANG; YUE; FARAWAY, 2018).

Na literatura estatística há várias formas para a escolha e a definição de prioris. Young e Smith (2005) indicaram que a adoção de prioris vagas, uniformes e impróprias sejam as formas

<sup>3</sup> Na literatura, é possível encontrar a notação de  $\pi(\boldsymbol{\theta} | \mathbf{y})$  para se referir a posteriori e  $\pi_0(\boldsymbol{\theta})$  para se referir a priori.

mais utilizadas na prática. Além destas, podem ser estabelecidas prioris subjetivas, em que há informações disponíveis e podem ser incorporadas no modelo, e prioris por conveniência, como o caso das conjugadas.

No caso de prioris conjugadas, as distribuições a priori e a posteriori pertencem a mesma família de distribuições, trazendo vantagens em termos de álgebra na explicitação da posteriori (WANG; YUE; FARAWAY, 2018). A abordagem é vista como uma forma conveniente e é comumente adotada em casos em que é possível estabelecê-la (YOUNG; SMITH, 2005).

Ao apresentar a análise Bayesiana dos GARMA(p, q), Andrade, Andrade e Ehlers (2015) adotaram densidades Gaussianas e com alta variabilidade para  $\boldsymbol{\theta}$ , tornando-as vagas e fazendo com que a informação proveniente da amostra seja dominante. Em aplicações de séries temporais, Broemeling (2019) utilizou diferentes prioris em modelos clássicos. Como exemplo, tem-se o uso de prioris Gaussianas em regressões com erros AR(1).

### 2.5.2 Métodos de Monte Carlo via cadeias de Markov

As distribuições  $p(\boldsymbol{\theta} | \mathbf{y})$  são determinadas após a especificação de  $p(\boldsymbol{\theta})$  e de  $L(\boldsymbol{\theta} | \mathbf{y})$ , mas o entendimento da informação atualizada nem sempre é trivial pois  $p(\boldsymbol{\theta} | \mathbf{y})$  podem ser multivariadas e de alta dimensão (WANG; YUE; FARAWAY, 2018). O que se faz é analisar a distribuição a posteriori marginal de cada parâmetro  $\theta_i$  que compõe  $\boldsymbol{\theta}$ , obtida da seguinte forma:

$$p(\theta_i | \mathbf{y}) = \int_{-\theta_i} p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-i}.$$

Essa integração pode ser complexa e soluções exatas são comumente restritas à classe de famílias conjugadas, sendo necessário utilizar métodos numéricos para aproximar estes resultados (WANG; YUE; FARAWAY, 2018).

Há diversos métodos para lidar com este problema. Na inferência Bayesiana, os métodos de Monte Carlo tornaram-se os mais utilizados, sendo baseados na geração de números pseudoaleatórios para simular amostras da posteriori de interesse (YOUNG; SMITH, 2005) e eficientes em problemas de alta dimensão (HASTINGS, 1970).

Dois métodos Monte Carlo via cadeias de Markov (MCMC) utilizados na literatura são denominados como amostrador *Gibbs* e algoritmo *Metropolis Hastings* (MH) e são fundamentados em uma simulação estocástica cuja a distribuição estacionária é a posteriori. Para compreender essas metodologias é necessário definir uma cadeia de Markov.

**Definição 1** (Korn, Korn e Kroisandt (2010)) *Um processo estocástico  $\{X(n), n \in \mathbb{N}\}$  é uma cadeia de Markov se  $P(X(n+1) = j | X(n) = i, X(n-1) = i_{n-1}, \dots, X(0) = i_0) = P(X(n+1) = j | X(n) = i) =: p_{ij}(n)$ ,  $\forall i, j, i_k \in S$ , em que  $S$  é denominado espaço de estados e  $p_{ij}(n)$  é denominada matriz de transição do tempo  $n$  para o tempo  $n+1$ .*

Os métodos MCMC requerem algumas propriedades desta cadeia, de modo que ela seja homogênea, irredutível e aperiódica, garantindo que a distribuição de  $X(n)$  convirja para a distribuição estacionária. Detalhes teóricos sobre cadeias de Markov podem ser vistos em Korn, Korn e Kroisandt (2010).

A ideia do amostrador de *Gibbs* é construir uma cadeia de Markov em que o núcleo de transição é composto pelas distribuições condicionais completas de  $\theta$  (NOGUEIRA, 2004). Em casos em que não é possível estabelecê-las ou identificá-las utiliza-se o MH. A implementação simultânea dos dois procedimentos também pode ser efetuada caso a condicional de um componente de  $\theta$  seja identificada e as demais não (YOUNG; SMITH, 2005).

Conforme Korn, Korn e Kroisandt (2010), o MH constrói uma cadeia de Markov iniciada a partir de um estado arbitrário e com probabilidade de transição  $p_{i,j}$ , que representa a probabilidade de movimento do estado  $i$  para o  $j$ , a fim de que a cadeia seja reversível, ou seja:

$$p(i | \mathbf{y})p_{i,j} = p(j | \mathbf{y})p_{j,i}.$$

De uma forma geral, o algoritmo MH é dado como segue:

1. Defina o núcleo de transição  $q(i, j)$ , o estado inicial  $X(0)$  e o número de iterações  $N$ .
2. Para  $k = 0$  até  $N - 1$ , faça:
  - a) gere  $\theta^*$  de  $q(X(k), \cdot)$  e um valor  $u \sim U[0, 1]$ ;
  - b) calcule a probabilidade de aceitação  $\alpha(X(k), \theta^*) = \min \left\{ 1, \frac{p(\theta^* | \mathbf{y})q(\theta^*; X(k))}{p(X(k) | \mathbf{y})q(X(k); \theta^*)} \right\}$ ;
  - c) se  $\alpha(X(k), \theta^*) > u$ , então  $X(k+1) = \theta^*$ ,  $k = k + 1$  e retorne em (a). Senão, retorne em (a).

A probabilidade de aceitação pode ser resumida em  $\alpha(X(k), \theta^*) = \min \left\{ 1, \frac{p(\theta^* | \mathbf{y})}{p(X(k) | \mathbf{y})} \right\}$ , caso  $q(i, j)$  seja simétrica, ou seja,  $q(X(k); \theta^*) = q(\theta^*; X(k))$  (KORN; KORN; KROISANDT, 2010). Há várias formas para a escolha do núcleo de transição: Hastings (1970) indicou que núcleos com alta rejeição devem ser evitados e Korn, Korn e Kroisandt (2010) sugeriram que uma possibilidade é utilizar um núcleo Gaussiano centrado em  $X(k)$ .

Como este processo é dinâmico, a cadeia atingirá um comportamento estacionário ao longo do tempo. Amostras anteriores ao estado de equilíbrio devem ser descartadas, o que

se denomina período *burn-in*<sup>4</sup>, e dada a correlação presente entre as amostras é usual utilizar uma amostragem sistemática, mantendo um espaçamento entre elas, o que se denomina como período *thin* (KORN; KORN; KROISANDT, 2010).

Uma metodologia alternativa aos procedimentos *Gibbs* e MH tem sido utilizada na inferência Bayesiana e é denominada como Monte Carlo Hamiltoniano (HMC), sendo um método MCMC baseado na dinâmica Hamiltoniana para a construção de cadeias de Markov (OLIVEIRA, 2021), possibilitando um movimento mais rápido pela distribuição alvo (GELMAN *et al.*, 2014).

### 2.5.2.1 Monte Carlo Hamiltoniano

O amostrador HMC, também conhecido como Monte Carlo Híbrido, foi proposto em Duane *et al.* (1987) e é fundamentado em uma teoria da física em que se descreve a energia total de um sistema de partículas através de uma função chamada Hamiltoniana (XAVIER, 2019).

Na função Hamiltoniana, considera-se um vetor posição  $\boldsymbol{\theta} \in \mathbb{R}^d$  e um vetor momento  $\mathbf{p} \in \mathbb{R}^d$ . A energia total do sistema é descrita da seguinte forma:

$$H(\boldsymbol{\theta}, \mathbf{p}) = U(\boldsymbol{\theta}) + K(\mathbf{p}),$$

sendo  $U(\boldsymbol{\theta})$  e  $K(\mathbf{p})$  a energia potencial e cinética, respectivamente. No caso do HMC,  $U(\boldsymbol{\theta})$  representa o negativo do logaritmo da densidade a posteriori, ou seja,  $U(\boldsymbol{\theta}) = -\log(p(\boldsymbol{\theta} | \mathbf{y}))$  e  $K(\mathbf{p}) = \frac{1}{2}\mathbf{p}^\top \mathbf{M}^{-1}\mathbf{p}$ . Em que  $\mathbf{M}$  é uma matriz simétrica, positiva definida e usualmente diagonal (TORRES, 2018).

Com a definição de  $H(\boldsymbol{\theta}, \mathbf{p})$ , são utilizadas equações de movimento para determinar como  $\boldsymbol{\theta}$  e  $\mathbf{p}$  se movem ao longo do tempo  $t$  através das seguintes derivadas parciais:

$$\begin{aligned} \frac{\partial \boldsymbol{\theta}}{\partial t} &= \frac{\partial H}{\partial \mathbf{p}} \\ \frac{\partial \mathbf{p}}{\partial t} &= -\frac{\partial H}{\partial \boldsymbol{\theta}}, \end{aligned} \tag{2.12}$$

que são responsáveis por mapear o estado do processo do instante  $t$  para o instante  $t + s$  (NEAL, 2011).

---

<sup>4</sup> Alguns estudos e programas computacionais têm denominado esse período como *warm-up*.

As soluções das equações (2.12) geralmente não podem ser obtidas analiticamente e são aproximadas por métodos numéricos como o de Euler, Euler modificado e o de Störmer-Verlet, também conhecido como *leapfrog* (XAVIER, 2019). A aproximação é dada pela discretização do tempo utilizando pequenos passos ( $\varepsilon$ ) a partir do tempo zero (NEAL, 2011). O método *leapfrog* produz bons resultados e é um dos mais populares, dado por:

$$\begin{aligned}\mathbf{p}\left(t + \frac{\varepsilon}{2}\right) &= \mathbf{p}(t) + \left(\frac{\varepsilon}{2}\right) \nabla_{\boldsymbol{\theta}} \log(p(\boldsymbol{\theta}(t) | \mathbf{y})) \\ \boldsymbol{\theta}(t + \varepsilon) &= \boldsymbol{\theta}(t) + \varepsilon \mathbf{M}^{-1} \mathbf{p}\left(t + \frac{\varepsilon}{2}\right) \\ \mathbf{p}(t + \varepsilon) &= \mathbf{p}\left(t + \frac{\varepsilon}{2}\right) + \left(\frac{\varepsilon}{2}\right) \nabla_{\boldsymbol{\theta}} \log(p(\boldsymbol{\theta}(t + \varepsilon) | \mathbf{y})),\end{aligned}$$

para um  $\varepsilon > 0$ , com erro local de ordem  $\varepsilon^3$  e global de ordem  $\varepsilon^2$  (NEAL, 2011; XAVIER, 2019).

A ligação entre essa dinâmica e a Bayesiana é feita ao se considerar  $\boldsymbol{\theta}$  como parâmetros de interesse e  $\mathbf{p}$  como auxiliares, de modo que  $\mathbf{p} \sim N_d(\boldsymbol{\theta}, \mathbf{M})$  (XAVIER, 2019). Segundo Gelman *et al.* (2014), isso faz com que a posteriori  $p(\boldsymbol{\theta} | \mathbf{y})$  seja aumentada resultando em  $p(\boldsymbol{\theta}, \mathbf{p} | \mathbf{y}) = p(\boldsymbol{\theta} | \mathbf{y})p(\mathbf{p})$ . Desta forma, segundo Xavier (2019), considera-se que o negativo do logaritmo da conjunta de  $(\boldsymbol{\theta}, \mathbf{p})$  é a função Hamiltoniana, ou seja:

$$H(\boldsymbol{\theta}, \mathbf{p}) = -\log(p(\boldsymbol{\theta} | \mathbf{y})) + \frac{1}{2} \log\left((2\pi)^d | \mathbf{M} | \right) + \frac{1}{2} \mathbf{p}^\top \mathbf{M}^{-1} \mathbf{p}.$$

O objetivo do HMC é suprir o comportamento de passeio aleatório local observado no *Gibbs* e no MH, em que a simulação pode tornar-se ineficiente e levar um longo período de tempo (GELMAN *et al.*, 2014). Entretanto, o HMC não é ‘milagroso’ pois a amostragem de algumas distribuições a posteriori pode ser uma tarefa complicada para qualquer algoritmo (MCELREATH, 2019).

Conforme Gelman *et al.* (2014), o HMC funciona da seguinte forma:

1. Cada iteração é iniciada atualizando  $\mathbf{p}$  a partir de sua respectiva posteriori, especificada como normal.
2. Procede-se com a atualização de  $(\boldsymbol{\theta}, \mathbf{p})$  com  $L$  passos *leapfrog* para um dado  $\varepsilon$ . Em cada passo  $L$ :

a) utiliza-se o gradiente de  $\log(p(\boldsymbol{\theta} | \mathbf{y}))$  para fazer um meio passo de  $\mathbf{p}$ , ou seja:

$$\mathbf{p} = \mathbf{p} + \frac{\varepsilon}{2} \frac{\partial \log(p(\boldsymbol{\theta} | \mathbf{y}))}{\partial \boldsymbol{\theta}};$$

b) utiliza-se  $\mathbf{p}$  para atualizar  $\boldsymbol{\theta}$  da seguinte forma:

$$\boldsymbol{\theta} = \boldsymbol{\theta} + \varepsilon \mathbf{M}^{-1} \mathbf{p};$$

c) realiza-se novamente um meio passo em  $\mathbf{p}$ :

$$\mathbf{p} = \mathbf{p} + \frac{\varepsilon}{2} \frac{\partial \log(p(\boldsymbol{\theta} | \mathbf{y}))}{\partial \boldsymbol{\theta}}.$$

3. Denomina-se como  $(\boldsymbol{\theta}^{t-1}, \mathbf{p}^{t-1})$  o vetor de valores prévios ao processo de atualização de *leapfrog* e  $(\boldsymbol{\theta}^*, \mathbf{p}^*)$  como os valores após os  $L$  passos.

4. Avalie a aceitação ou rejeição dos valores candidatos:

$$r = \frac{p(\boldsymbol{\theta}^* | \mathbf{y})p(\mathbf{p}^*)}{p(\boldsymbol{\theta}^{t-1} | \mathbf{y})p(\mathbf{p}^{t-1})}.$$

5. Defina  $\boldsymbol{\theta}^t$  como:

$$\boldsymbol{\theta}^t = \begin{cases} \boldsymbol{\theta}^*, & \text{com probabilidade } \min(r, 1) \\ \boldsymbol{\theta}^{t-1} & \text{caso contrário.} \end{cases}$$

6. Assim como nos demais algoritmos MCMC, repita as iterações até a convergência e a avalie.

Observa-se, então, que a dinâmica é utilizada para propor candidatos ao novo estado da cadeia. Isso é feito ao se construir  $L$  passos *leapfrog* ao redor da função  $H(\boldsymbol{\theta}, \mathbf{p})$  e considerar o último estado como candidato (OLIVEIRA, 2021). Em termos de construção, recomenda-se a criação de múltiplas cadeias iniciadas arbitrariamente, com descarte do período inicial e avaliação de convergência através do Fator de redução de escala potencial ( $\hat{R}$ ) (STAN, 2022).

O HMC pode ter um custo computacional mais elevado quando comparado com os algoritmos *Gibbs* e MH. Porém, as propostas são mais eficientes, não sendo necessário um elevado número de amostras para se descrever a posteriori, apresentando melhor desempenho em situações com muitos parâmetros (MCELREATH, 2019), baixa correlação entre as amostras e alta taxa de aceitação (BURDA; MAHEU, 2013).

Os valores de  $L$ ,  $\varepsilon$  e  $\mathbf{M}$  podem impactar o desempenho do algoritmo mas podem ser calibrados durante o período *warm-up* (HOFFMAN; GELMAN, 2014; CONCEIÇÃO; SUZUKI;

ANDRADE, 2021). Uma forma de adaptação de  $\mathbf{M}$  é utilizar o inverso da matriz de covariâncias de  $p(\boldsymbol{\theta} | \mathbf{y})$ , aumentando a eficiência do algoritmo (GELMAN *et al.*, 2014), ou realizar a estimação de um intervalo ótimo de discretização ( $\epsilon$ ) (STAN, 2022).

Como uma forma de extensão ao HMC, o *No-U-Turn Sampler* (NUTS) foi proposto por Hoffman e Gelman (2014) utilizando uma forma recursiva para selecionar estados candidatos e realizando uma parada automática no momento em que o algoritmo começa refazer os passos. O que elimina a necessidade de que  $L$  seja fixo em todas as iterações, sendo adaptado em cada uma delas. Detalhes sobre o NUTS estão em Hoffman e Gelman (2014), Oliveira (2021) e Stan (2022).

Os algoritmos HMC e NUTS estão implementados no R Core Team (2022) através do pacote *Rstan* (STAN DEVELOPMENT TEAM, 2018). Em linguagem *Python*, o pacote é denominado *PyStan* (PYSTAN, 2019). A documentação e a implementação de modelos, incluindo exemplos de aplicação, encontram-se em Stan (2022).

### 2.5.2.2 Monitoramento da convergência

Há vários métodos para se analisar a convergência das cadeias de Markov. Nogueira, Sáfadi e Ferreira (2004) sugeriram que sejam usadas combinações desses métodos. Alguns critérios para tal propósito estão apresentados a seguir, mais especificamente, as formas propostas por Heidelberger e Welch (1983), Geweke (1991), Gelman e Rubin (1992) e Raftery e Lewis (1995).

#### Heidelberger e Welch (1983)

É um critério baseado na densidade espectral e utiliza a teoria *Browniana* para testar a hipótese nula de estacionariedade (NOGUEIRA; SÁFADI; FERREIRA, 2004). Primeiramente, o teste verifica se a sequência completa é estacionária. Caso a hipótese nula seja rejeitada descarta-se 10% das iterações iniciais e isso pode ser repetido até o limite de 50% das iterações.

Desta forma, assume-se que a sequência é não estacionária se o teste falhar em 6 resultados consecutivos (HEIDELBERGER; WELCH, 1983). Caso ocorram falhas nas primeiras iterações, o número de iterações descartadas pelo teste pode ser utilizado como período *burn-in* (NOGUEIRA; SÁFADI; FERREIRA, 2004).

O procedimento também realiza o teste de *half-width* na sequência que passou no teste de estacionariedade para verificar se a média está sendo calculada com uma acurácia  $\lambda$  (NO-



GUEIRA; SÁFADI; FERREIRA, 2004). É construído um intervalo de confiança a 95% e verifica-se se  $\frac{L_s - L_i}{2\theta} \leq \lambda$ , sendo  $L_i$  e  $L_s$  o limite inferior e superior do intervalo (HEIDELBERGER; WELCH, 1983). Caso isso ocorra, a média está sendo estimada com uma acurácia satisfatória.

### **Geweke (1991)**

É um teste para analisar a convergência a partir da comparação das médias nos  $m_1$  passos iniciais e  $m_2$  passos finais da cadeia. Sob a hipótese nula de convergência, adota-se a seguinte estatística de teste:

$$G = \frac{\bar{\theta}_1 - \bar{\theta}_2}{\sqrt{\frac{\hat{S}_1(0)}{m_1} + \frac{\hat{S}_2(0)}{m_2}}} \sim N(0, 1).$$

Sendo  $\bar{\theta}_i$  a média amostral e a densidade espectral estimada para as  $m_i$  iterações aplicada na frequência 0,  $\hat{S}_i(\cdot)$ ,  $\{i = 1, 2\}$ . É usual considerar  $m_1 = 10\%$  dos passos iniciais e  $m_2 = 50\%$  dos finais.

A ideia do procedimento é que caso as amostras sejam provenientes da distribuição estacionária, a hipótese de igualdade de  $\theta_1$  e  $\theta_2$  não é rejeitada para um dado nível  $\alpha$ . O método fornece ao pesquisador se a convergência não foi alcançada (NOGUEIRA; SÁFADI; FERREIRA, 2004) e é interessante que este seja realizado após o descarte das amostras iniciais (GEWEKE, 1991).

### **Gelman e Rubin (1992)**

Utiliza sequências de mesmo comprimento para tomar a decisão a respeito da convergência. A ideia é avaliar se as sequências apresentam comportamento similar após o *burn-in* (NOGUEIRA; SÁFADI; FERREIRA, 2004). O diagnóstico ficou conhecido como Fator de redução de escala potencial,  $\hat{R}$ , mas pode apresentar falhas em situações de caudas pesadas (VEHTARI *et al.*, 2021). Uma ampliação do  $\hat{R}$  foi dada em Gelman *et al.* (2014) e denominada como *split- $\hat{R}$* .

Para a construção do *split- $\hat{R}$* , cada sequência é dividida ao meio após o descarte inicial e considera-se como  $m$  o número de sequências após as divisões e  $n$  o tamanho de cada uma (GELMAN *et al.*, 2014). Por exemplo, se foram construídas duas cadeias de tamanho 1000 com descarte das 500 amostras iniciais,  $m = 4$  e  $n = 250$ . A partir disso, estima-se a variância

entre (B) e dentro (W) de cada sequência utilizando os seguintes estimadores:

$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\theta}_{.j} - \bar{\theta}_{..})^2,$$

$$W = \frac{1}{m} \sum_{j=1}^m s_j^2,$$

sendo  $\bar{\theta}_{.j} = \frac{1}{n} \sum_{i=1}^n \theta_{ij}$ ,  $\bar{\theta}_{..} = \frac{1}{m} \sum_{j=1}^m \bar{\theta}_{.j}$  e  $s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\theta_{ij} - \bar{\theta}_{.j})^2$  para  $i = \{1, \dots, n\}$  e  $j = \{1, \dots, m\}$  e estima-se a variância marginal a posteriori,  $var(\theta | \mathbf{y})$ , ponderada por B e W, isto é:

$$\widehat{var}(\theta | \mathbf{y}) = \frac{n-1}{n} W + \frac{B}{n},$$

e então o *split- $\hat{R}$*  é estimado por:

$$\hat{R} = \sqrt{\frac{\widehat{var}(\theta | \mathbf{y})}{W}},$$

possibilitando monitorar a convergência (GELMAN *et al.*, 2014). O *split- $\hat{R}$*  é um dos critérios utilizados pelo Stan (2022), com valor esperado igual a um e sendo desejado valores menores que 1,05.

### **Raftery e Lewis (1995)**

É um critério que fornece uma estimativa do número de iterações a serem eliminadas inicialmente, do *thin* e do número de amostras necessárias para a convergência. Em problemas multivariados, o procedimento retorna resultados para cada variável e é necessário escolher o máximo entre estes (NOGUEIRA; SÁFADI; FERREIRA, 2004).

A metodologia também estima o fator de dependência, entendido como o acréscimo multiplicativo de iterações para se atingir a distribuição estacionária (NOGUEIRA; SÁFADI; FERREIRA, 2004). Fatores de dependência maiores que um indicam elevada dependência entre as amostras e maiores que cinco sugerem que a convergência não foi atingida (RAFTERY; LEWIS, 1995).

Ao se utilizar métodos MCMC, o estudo de Nogueira, Sáfadi e Ferreira (2004) recomendou planejar um modelo piloto para a determinação do tamanho das cadeias via critério de Raftery e Lewis (1995) e monitorar a convergência na proximidade da especificação sugerida através de outros diagnósticos.

### 2.5.3 Estimação pontual e intervalar

Com a obtenção da marginal a posteriori, é usual sumariá-la a partir de estatísticas como a média, a moda e a mediana. Também pode ser construído o gráfico da densidade marginal, possibilitando avaliar comportamentos assimétricos e escolher um estimador pontual que resuma a informação atualizada (WANG; YUE; FARAWAY, 2018).

De forma intervalar, podem ser construídos intervalos de credibilidade. Métodos para cálculos de intervalo de credibilidade não são únicos, mas no geral tem-se maior interesse em intervalos de menor amplitude (WANG; YUE; FARAWAY, 2018). Conforme Young e Smith (2005), deseja-se encontrar uma região  $R$  que maximiza a probabilidade a posteriori de  $\theta \in R$ , denominada de *Highest posterior density* (HPD) ou de alta densidade.

**Definição 2** (Box e Tiao (1992)) *Uma região  $R \subseteq \Omega$  é uma região HPD com probabilidade  $(1 - \alpha)100\%$ , se:*

1.  $P_r(\theta \in R | \mathbf{y}) = 1 - \alpha$ ;
2. Para  $\theta_1 \in R$  e  $\theta_2 \notin R$  a  $p(\theta_1 | \mathbf{y}) \geq p(\theta_2 | \mathbf{y})$ .

Intervalos HPD nem sempre são utilizados em situações em que é complicado se obter a densidade a posteriori. Uma alternativa para densidades a posteriori aproximadamente Gaussianas é construir um intervalo aproximado baseado na normal. Em casos de densidades que não são unimodais pode-se estabelecer a união de intervalos, formando os conjuntos HPD (YOUNG; SMITH, 2005).

### 2.5.4 Comparação e adequação de modelos

Há diversas metodologias disponíveis que possibilitam a comparação e a análise da adequação de modelos. Nesta seção estão apresentadas duas delas: o Critério de Informação da Deviância (DIC) e a Ordenada Preditiva Condicional (CPO), que é computada baseada na distribuição preditiva.

#### 2.5.4.1 Critério de informação da deviância

O DIC é uma medida que permite comparar modelos candidatos, onde modelos com menores valores de DIC são preferíveis (WANG; YUE; FARAWAY, 2018; ASSIS, 2020). Inicialmente define-se a deviância da seguinte forma:

$$D(\boldsymbol{\theta}) = -2\log(p(\mathbf{y} | \boldsymbol{\theta})).$$

Como  $D(\boldsymbol{\theta})$  é uma variável aleatória, utiliza-se a deviança média a posteriori,  $E(D(\boldsymbol{\theta}))$ , denominada como  $\bar{D}$ . Com isso, o DIC é estimado por:

$$DIC = \bar{D} + p_D,$$

sendo  $p_D = \bar{D} - D(\bar{\boldsymbol{\theta}})$  uma penalidade do número efetivo de parâmetros.

#### 2.5.4.2 Distribuição preditiva e a ordenada preditiva condicional

Tratando-se de séries temporais, a seleção de modelos pode ser baseada em uma medida de erros de previsão caso este seja o objetivo do pesquisador. Para isso, é necessário construir a distribuição condicional das futuras observações a partir da posteriori, denominada distribuição preditiva, dada por:

$$G(y^* | \mathbf{y}) = \int p(y^* | \boldsymbol{\theta})p(\boldsymbol{\theta} | \mathbf{y})d\boldsymbol{\theta},$$

sendo  $y^*$  a nova observação (YOUNG; SMITH, 2005). A estimativa pontual de  $y^*$  é dada pela média, mediana ou alguma outra função perda de  $G(\cdot)$ .

Em certas situações a preditiva não tem forma fechada. Com isso, é necessário utilizar uma aproximação Monte Carlo para estabelecer o valor predito de  $y^*$ . Nos modelos GARMA(p, q), por exemplo, Andrade, Andrade e Ehlers (2015) utilizaram o valor esperado da preditiva via simulação MCMC, possibilitando a construção de intervalos HPD para os valores preditos.

A partir da preditiva pode-se obter a Ordenada preditiva condicional, do inglês, *Conditional predictive ordinate* (CPO). A CPO fornece a evidência da verossimilhança para cada observação  $y_i$  que compõe a amostra e é calculada por  $G(y_i | \mathbf{y}_{(-i)})$ , em que  $\mathbf{y}_{(-i)}$  representa o vetor de observações com exceção da  $i$ -ésima. Valores baixos de CPO indicam observações que estão mal ajustadas (TURKMAN; PAULINO; MÜLLER, 2019).

Considerando a soma da CPO calculada para cada observação  $y_i$  pertencente à amostra obtém-se a chamada *Log pseudo-marginal likelihood* (LPML), dada por  $LPML = \sum_{i=1}^n \log(CPO_i)$ , indicando preferência para modelos com maiores valores de LPML (TURKMAN; PAULINO; MÜLLER, 2019) e sendo utilizada como uma alternativa ao critério DIC (WANG; YUE; FARAWAY, 2018).

### 3 CONCLUSÃO

Entende-se que o ponto central deste trabalho foi abordar modelos para séries temporais de contagem realizando flexibilizações e ampliações para algumas distribuições de probabilidade discretas e estruturas de dependência temporal. O desenvolvimento do trabalho foi dado em três ensaios.

No primeiro ensaio foram exploradas as distribuições Binomial negativa, Poisson e Poisson inversa Gaussiana com a inserção da estrutura AR(p) ao modelar dados de furtos de veículos. A análise possibilita a disseminação de modelos que comportem particularidades como a alta dispersão, contribuindo também com a flexibilização para outras distribuições e propiciando alternativas aos tradicionais modelos Gaussianos.

Ao considerar características como a alta dispersão e excesso de zeros, o segundo ensaio abordou as versões zero-ajustadas das distribuições Binomial negativa, Poisson e Poisson inversa Gaussiana com a inserção da estrutura AR(p) para comportar a correlação serial. Foi realizada uma aplicação em dados de mortalidade em decorrência da dengue, possibilitando avaliar tendências do nível de mortalidade em períodos futuros.

No terceiro ensaio adotou-se um modelo com distribuição Poisson zero-ajustada cujos parâmetros variam no tempo. Isso possibilita analisar séries com excesso de zeros sob a suposição de que a probabilidade de ocorrência de zeros é dependente do tempo, permitindo realizar previsões das contagens e das probabilidades futuras e podendo ser ampliada para outras distribuições zero-ajustadas como a Binomial negativa.

Por fim, ressalta-se que os passos algébricos e os resultados de simulação obtidos em cada um dos ensaios desta tese podem e serão fornecidos a interessados mediante solicitação aos autores. Além disso, no Apêndice A, tem-se exemplos em R dos códigos utilizados em cada ensaio, podendo ser ampliados para outras estruturas de dependência temporal.

## REFERÊNCIAS

- ALBARRACIN, O.; ALENCAR, A.; HO, L. Generalized autoregressive and moving average models: multicollinearity, interpretation and a new modified model. **Journal of statistical computation and simulation**, v. 89, n. 10, p. 1819-1840, 2019.
- ALQAWBA, M.; DIAWARA, N.; CHAGANTY, N. Zero-inflated count time series models using Gaussian copula. **Sequential analysis**, v. 38, n. 3, p. 342–357, 2019.
- ANDRADE, B.; ANDRADE, M.; EHLERS, R. Bayesian Gamma models for count data. **Communications in statistics: case studies, data analysis and applications**, v. 1, n. 4, p. 192–205, 2015.
- ASSIS, C. **Modelos Bayesianos zero-modificados para séries temporais de contagem**. 2020. 111 p. Tese (Doutorado em Estatística) - Universidade de São Paulo, São Carlos, 2020.
- BARRETO-SOUZA, W. Mixed Poisson INAR(1) processes. **Statistical papers**, v. 60, p. 2119–2139, 2017.
- BAYER, F.; CINTRA, R.; CRIBARI-NETO, F. Beta seasonal autoregressive moving average models. **Journal of statistical computation and simulation**, v. 88, n. 15, p. 2961–2981, 2018.
- BENJAMIN, M.; RIGBY, R.; STASINOPOULOS, M. Generalized autoregressive moving average models. **Journal of the American statistical association**, v. 461, n. 98, p. 214–223, 2003.
- BOX, G.; PIERCE, D. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American statistical association**, v. 65, n. 332, p. 1509–1526, 1970.
- BOX, G.; JENKINS, G. **Time series analysis, forecasting and control**. Califórnia: Holden-Day, 1976.
- BOX, G.; TIAO, G. **Bayesian inference in statistical analysis**. New York: John Wiley & Sons, 1992.
- BRIËT, O.; AMERASINGHE, P.; VOUNATSOU, P. Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. **Plos one**, v. 8, n. 6, p. 1–9, 2013.
- BROEMELING, L. **Bayesian analysis of time series**. Boca Raton: CRC Press, 2019.
- BURDA, M.; MAHEU, J. Bayesian adaptive Hamiltonian Monte Carlo with an application to high-dimensional BEKK GARCH models. **Studies in nonlinear dynamics and econometrics**, v. 17, n. 4, p. 345-372, 2013.
- CONCEIÇÃO, K.; SUZUKI, A.; ANDRADE, M. A Bayesian approach for zero-modified Skellam model with Hamiltonian MCMC. **Statistical methods & applications**, v. 30, n. 2, p. 747-765, 2021.
- DAVIS, R.; DUNSMUIR, W.; STREETT, S. Observation-driven model for Poisson counts. **Biometrika**, v. 90, n. 4, p. 777–790, 2003.

- DAVIS, R. *et al.* Count time series: a methodological review. **Journal of the American statistical association**, v. 116, n. 535, p. 1533-1547, 2021.
- DUANE, S. *et al.* Hybrid Monte Carlo. **Physics letters B**, v. 195, n. 2, p. 216-222, 1987.
- DUNN, P.; SMYTH, G. Randomized quantile residuals. **Journal of computational and graphical statistics**, v. 5, n. 3, p. 236-244, 1996.
- DUNSMUIR, W. Generalized linear autoregressive moving average models. *In*: DAVIS, R. *et al.* **Handbook of discrete-valued time series**. Boca Raton: CRC Press, 2016.
- EHLERS, R. A **Conway-Maxwell-Poisson Gamma model for count data**. 2019. Disponível em: <https://arxiv.org/abs/1901.07473>. Acesso em: 10 nov. 2022.
- FENG, C. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. **Journal of statistical distributions and applications**, v. 8, n. 1, p. 1-19, 2021.
- FRANCO, G.; MIGON, H.; MARCOS, P. A **Monte Carlo study and data analysis for time series of counts**. 2015. Disponível em: [http://www.est.ufmg.br/portal/arquivos/rts/RTP\\_01\\_2015.pdf](http://www.est.ufmg.br/portal/arquivos/rts/RTP_01_2015.pdf). Acesso em: 11 jun. 2021.
- GELMAN, A.; RUBIN, D. Inference from iterative simulation using multiple sequences. **Statistical science**, v. 7, n. 4, p. 457-472, 1992.
- GELMAN, A. *et al.* **Bayesian data analysis**. 3. ed. Boca Raton: CRC Press, 2014.
- GEWEKE, J. **Evaluating the accuracy of sampling based approaches to the calculation of posterior moments**. 1991. Disponível em: <https://ideas.repec.org/p/fip/fedmsr/148.html>. Acesso em: 06 jan. 2020.
- GHAHRAMANI, M.; WHITE, S. Time series regression for zero-inflated and overdispersed count data: a functional response model approach. **Journal of statistical theory and practice**, v. 14, n. 29, p. 1-18, 2020.
- GONÇALVES, J.; BARRETO-SOUZA, W. Flexible regression models for counts with high-inflation of zeros. **Metron**, v. 78, n. 1, p. 71-95, 2020.
- HASHIM, L.; HASHIM, K.; SHIKER, M. An application comparison of two poisson models on zero count data. **Journal of physics**, v. 1818, p. 1-12, 2021.
- HASTINGS, W. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, n. 1, p. 97-109, 1970.
- HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations research**, v. 31, n. 6, p. 1109-1144, 1983.
- HELLER, G.; COUTURIER, D.; HERITIER, S. Beyond mean modelling: bias due to misspecification of dispersion in Poisson inverse Gaussian regression. **Biometrical journal**, v. 61, n. 2, p. 333-342, 2018.
- HILBE, J. **Modeling count data**. New York: Cambridge University Press, 2014.
- HOFFMAN, M.; GELMAN, A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. **Journal of machine learning research**, v. 15, n. 1, p. 1351-1381, 2014.

- KORN, R.; KORN, E.; KROISANDT, G. **Monte Carlo methods and models in finance and insurance**. London: CRC Press, 2010.
- LIMA, T. **Modelos INAR e RCINAR, estimação e aplicação**. 2013. 107 p. Dissertação (Mestrado em Estatística) - Universidade de São Paulo, São Paulo, 2013.
- LJUNG, G.; BOX, G. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297–303, 1978.
- MCELREATH, R. **Statistical rethinking: a Bayesian course with examples in R and Stan**. 2. ed. [S.l.]: CRC Press, 2019.
- MELO, M.; ALENCAR, A. Conway-Maxwell–Poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data. **Journal of time series analysis**, v. 41, n. 6, p. 830-857, 2020.
- MORETTIN, P.; TOLOI, C. **Análise de séries temporais**. 2. ed. São Paulo: Blucher, 2006.
- MORETTIN, P. **Econometria financeira: um curso em séries temporais financeiras**. São Paulo: Blucher, 2017.
- MULLAHY, J. Specification and testing of some modified count data models. **Journal of econometrics**, v. 33, n. 3, p. 341–365, 1986.
- NEAL, R. MCMC using Hamiltonian dynamics. *In*: BROOKS, S. *et al.* **Handbook of Markov chain Monte Carlo**. [S.l.]: Chapman & Hall, 2011.
- NELDER, J.; WEDDERBURN, R. Generalized linear models. **Journal of the royal statistical society**, v. 135, n. 3, p. 370-384, 1972.
- NOGUEIRA, D. **Proposta e avaliação de critérios de convergência para o método de Monte Carlo via cadeias de Markov: casos uni e multivariados**. 2004. 121 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2004.
- NOGUEIRA, D.; SÁFADI, T.; FERREIRA, D. Avaliação de critérios de convergência para o método de Monte Carlo via cadeias de Markov. **Revista Brasileira de estatística**, v. 65, n. 224, p. 59–88, 2004.
- OESSELMANN, C. **Equações de estimação generalizadas com resposta Binomial negativa: modelando dados correlacionados de contagem com sobredispersão**. 2016. 125 p. Dissertação (Mestrado em Estatística) - Universidade de São Paulo, São Paulo, 2016.
- OLIVEIRA, E. **Contribuições para modelos de diagnóstico cognitivo**. 2021. 146 p. Tese (Doutorado em Estatística), Universidade de São Paulo, São Carlos, 2021.
- OUMA, V.; MWALILI, S.; KIBERIA, A. Poisson inverse Gaussian (PIG) model for infectious disease count data. **American journal of theoretical and applied statistics**, v. 5, n. 5, p. 326–333, 2016.
- PAYNE, E. *et al.* Approaches for dealing with various sources of overdispersion in modeling count data: scale adjustment versus modeling. **Statistical methods in medical research**, v. 26, n. 4, p. 1802-1823, 2017.



- PYSTAN. **PyStan**. 2019. Disponível em: <https://pystan.readthedocs.io/en/latest/>. Acesso em: 08 fev. 2023.
- RAFTERY, A.; LEWIS, S. The number of iterations, convergence diagnostics and generic metropolis algorithms. **Practical Markov chain Monte Carlo**, v. 7, n. 98, p. 763–773, 1995.
- RAFTERY, A. *et al.* Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. **Bayesian statistics**, v. 8, p. 1-45, 2007.
- R CORE TEAM. **R: a language and environment for statistical computing**. 2022. Disponível em: <https://www.R-project.org/>. 2022. Acesso em: 25 ago. 2022.
- RIBGY, R. *et al.* **Distributions for modeling location, scale, and shape**. Boca Raton: Chapman & Hall, 2019.
- ROCHA, A.; CRIBARI-NETO, F. Beta autoregressive moving average models. **Test**, v. 18, n. 3, p. 529–545, 2008.
- SATHISH, V.; MUKHOPADHYAY, S.; TIWARI, R. Autoregressive and moving average models for zero-inflated count time series. **Statistica Neerlandica**, v. 76, n. 2, p. 190-218, 2021.
- SHEPHARD, N. **Generalized linear autoregressions**. 1995. Disponível em: <https://ora.ox.ac.uk/objects/uuid:a1883838-ef1b-4a23-8872-c0b17111d3bd>. Acesso em: 05 abr. 2023.
- SILVA, N.; PEREIRA, I.; SILVA, M. Forecasting in INAR(1) model. **Revstat**, v. 7, n. 1, p. 119–134, 2009.
- SILVA, R. **Generalized autoregressive neural network models**. 2020. Disponível em: <https://arxiv.org/abs/2002.05676>. Acesso em: 01 abr. 2022.
- STAN DEVELOPMENT TEAM. **RStan: the R interface to Stan**. 2018. Disponível em: <http://mc-stan.org/>. Acesso em: 01 jun. 2022.
- STAN. **Stan reference manual**. 2022. Disponível em: [https://mc-stan.org/docs/2\\_29/reference-manual-2\\_29.pdf](https://mc-stan.org/docs/2_29/reference-manual-2_29.pdf). Acesso em: 05 ago. 2022.
- STASINOPOULOS, M. *et al.* **Flexible regression and smoothing: using GAMLSS in R**. New York: CRC Press, 2017.
- STASINOPOULOS, M.; RIGBY, R. **Gamlss.dist: distributions for generalized additive models for location scale and shape**. 2020. Disponível em: <https://CRAN.R-project.org/package=gamlss.dist>. Acesso em: 20 jun. 2021.
- TAWIAH, K.; IDDRISU, W.; ASOSEGA, K. Zero-inflated time series modelling of COVID-19 deaths in Ghana. **Journal of environmental and public health**, p. 1-9, 2021.
- TORRES, M. **Determinação de propriedades térmicas em problemas de condução de calor por inferência Bayesiana com o método de Monte Carlo Hamiltoniano**. 2018. 87 p. Dissertação (Mestrado em Modelagem Computacional) - Universidade do Estado do Rio de Janeiro, 2018.

- TSAY, R. **An introduction to analysis of financial data with R**. New Jersey: John Wiley & Sons, 2013.
- TURKMAN, M.; PAULINO, C.; MÜLLER, P. **Computational Bayesian statistics: an introduction**. United Kingdom: Cambridge University Press, 2019.
- VEHTARI, A. *et al.* Rank-normalization, folding, and localization: an improved  $\hat{R}$  for assessing convergence of MCMC. **Bayesian analysis**, v. 16, n. 2, p. 667-718, 2021.
- WANG, X.; YUE, Y.; FARAWAY, J. **Bayesian regression modeling with INLA**. Boca Raton: Chapman & Hall, 2018.
- XAVIER, C. **Métodos de Monte Carlo Hamiltoniano aplicados em modelos GARCH**. 2019. 87 p. Tese (Doutorado em Estatística) - Universidade de São Paulo, São Carlos, 2019.
- YANG, M.; ZAMBA, G.; CAVANAUGH, J. Markov regression models for count time series with excess zeros: a partial likelihood approach. **Statistical methodology**, v. 14, p. 26-38, 2013.
- YOUNG, G.; SMITH, R. **Essentials of statistical inference**. New York: Cambridge University Press, 2005.
- ZHA, L.; LORD, D.; ZOU, Y. The Poisson inverse Gaussian (PIG) generalized linear regression model for analyzing motor vehicle crash data. **Journal of transportation safety & security**, v. 8, n. 1, p. 18–35, 2016.
- ZUO, G. *et al.* Generalized Poisson hurdle model for count data and its application in ear disease. **Entropy**, v. 23, n. 9, 2021.
- ZUUR, A. *et al.* **Mixed effects models and extensions in ecology with R**. New York: Springer, 2009.

## **SEGUNDA PARTE - ARTIGOS**

**ARTIGO 1 - *Analysis of count time series: a Bayesian GARMA( $p$ ,  $q$ ) approach***

Redigido conforme as normas da revista *Austrian journal of statistics* (versão em processo de editoração).



## Analysis of Count Time Series: A Bayesian GARMA(p, q) Approach

Luiz Otávio de O. Pala  Marcela de M. Carvalho  Thelma Sáfyadi 

Federal University of Lavras   Federal University of Lavras   Federal University of Lavras

### Abstract

Extensions of the Autoregressive Moving Average, ARMA(p, q), class for modeling non-Gaussian time series have been proposed in the literature in recent years, being applied in phenomena such as counts and rates. One of them is the Generalized Autoregressive Moving Average, GARMA(p, q), that is supported by the Generalized Linear Models theory and has been studied under the Bayesian perspective. This paper aimed to study models for time series of counts using the Poisson, Negative binomial and Poisson inverse Gaussian distributions, and adopting the Bayesian framework. To do so, we carried out a simulation study and, in addition, we showed a practical application and evaluation of these models by using a set of real data, corresponding to the number of vehicle thefts in Brazil.

*Keywords:* autoregressive moving average models, count data, generalized linear models, mcmc, time series.

## 1. Introduction

The literature on time series was pointed out by Cox, Gudmundsson, Lindgren, Bondesson, Harsaae, Laake, Juselius, and Lauritzen (1981) and Khandelwal, Adhikari, and Verm (2015) as one of the most active topics in statistics, being employed in economics, physics, and engineering, for example. In the time domain, the ARMA(p, q) class, proposed by Box and Jenkins (1976), is widely used in applied studies, being an efficient option to create forecasts (Silva 2020).

Despite the usefulness and efficiency of the ARMA(p, q) models, this class is considered more appropriate for fitting Gaussian data, assuming a linear structure of the correlation (Silva 2020). In financial or count time series, these assumptions can be strong, as they can have an asymmetric behavior and heavier tails than the Gaussian distribution.

According to Dunsmuir and Scott (2015); Davis, Fokianos, Holan, Joe, Livsey, Lund, Pipiras, and Ravishanker (2021); Kong and Lund (2022), there has been a recent interest in proposing and building models for non-Gaussian time series. This theme was discussed in Cox *et al.* (1981), who specified observation-driven models, in which the dependence structure is related to the past of the series, and parameter-driven models, in which the dependence follows a latent process. For further details about observation-driven and parameter-driven models,

2

see [Davis \*et al.\* \(2021\)](#).

Concerning observation-driven models, [Benjamin, Rigby, and Stasinopoulos \(2003\)](#) introduced the GARMA(p, q) class considering that the conditional distribution, given the past information, belongs to the exponential family, with the conditional mean related to a linear predictor that contains explanatory variables and the dependence structure in the ARMA(p, q) form.

A Bayesian analysis of the GARMA(p, q) class was presented and discussed in the study of [Andrade, Andrade, and Ehlers \(2015\)](#), who used conditional distributions such as Poisson, Binomial, and Negative binomial. The authors indicated contributions using this approach in terms of point estimation and the range of the credible intervals of the parameters when modeling count data.

When dealing with count time series, overdispersion is commonly observed ([Barreto-Souza 2017](#); [Gonçalves and Barreto-Souza 2020](#)), a peculiarity that makes the Poisson not suitable, since it is equidispersed. A solution for solving this problem is to consider mixed Poisson distributions like the Negative binomial (NB) and Poisson inverse Gaussian (PIG) ([Dean, Lawless, and Willmot 1989](#)). In addition to these, alternative approaches to model correlated counts using different conditional distributions, such as the Conway-Maxwell-Poisson and Bernoulli-Geometric, have arisen in literature ([Davis \*et al.\* 2021](#); [Sales, Alencar, and Ho 2022](#)).

The PIG distribution was introduced into the time series context by [Barreto-Souza \(2017\)](#) using the INteger-valued AutoRegressive (INAR) structure. However, the use of the PIG in regression models was proposed in [Dean \*et al.\* \(1989\)](#) for modeling insurance data, being considered an attractive distribution in the presence of heavy tails. For further information on modeling insurance data using the PIG regression, see [Willmot \(1987\)](#).

Our main goal in this paper is to study models for time series of counts using the Poisson, Negative binomial, and Poisson inverse Gaussian distributions, adopting the Bayesian framework used by [Andrade \*et al.\* \(2015\)](#) and extending it to the Poisson inverse Gaussian. As an application example, we analyze the performance of these models by forecasting the number of vehicle thefts in the region of *Campinas*, Brazil.

This paper is organized as follows: In Section 2 we described the models and Section 3 describes the Bayesian analysis. Section 4 focuses on computational and simulation results. In 5, we presented the real data application. Some remarks and topics for future research are discussed in Section 6.

## 2. Models

Let  $Y$  be a time series equally spaced and indexed in the time  $t$ , for  $t = \{1, \dots, n\}$ . The set of previous information until the instant  $t - 1$  is denoted by  $\mathcal{F}_{t-1}$ , and it is given by  $\mathcal{F}_{t-1} = \{y_1, y_2, \dots, y_{t-1}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_{t-1}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{t-1}\}$ , being  $\mathbf{x}_t$  a vector containing  $r$  explanatory variables  $\mathbf{x}_t = (x_{t1}, \dots, x_{tr})^\top$ , under the restriction that  $r < n$ .

The Autoregressive and Moving Average components that compose the dependence structure ( $\tau_t$ ) are represented by the vectors  $\boldsymbol{\Phi} = (\phi_1, \dots, \phi_p)^\top$  and  $\boldsymbol{\Theta} = (\theta_1, \dots, \theta_q)^\top$ , respectively, where  $p$  and  $q$  are the Autoregressive and Moving Average orders of an ARMA(p, q) process. The vector of coefficients, which is related to the  $r$  explanatory variables of  $\mathbf{x}_t$ , is expressed by  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_r)^\top$ .

Also consider that the likelihood function, denoted as  $L(\boldsymbol{\theta} | Y)$ , is given by the product of the conditional distributions, and it can be approximated as shown in Equation 1:

$$L(\boldsymbol{\theta} | Y) \approx \prod_{t=m+1}^n p(y_t | \mathcal{F}_{t-1}), \quad (1)$$

where  $m$  are the  $m = \max(p, q)$  first observations of  $Y$  and  $\boldsymbol{\theta}$  represents the set of parameters

of each model.

### 2.1. Poisson

In this model, we assume that the conditional distribution of each observation, given the past information, follows a Poisson distribution. So,  $p(y_t | \mathcal{F}_{t-1})$  is given by Equation 2:

$$p(y_t | \mathcal{F}_{t-1}) = \exp \{y_t \log(\mu_t) - \mu_t - \log(y_t!)\}, \quad (2)$$

being  $\Omega = \{\mu_t | \mu_t > 0\}$  with  $y_t \in \mathbb{N}$ . As known, the Poisson belongs to the exponential family with logarithm as the canonical link function, being it a submodel of the GARMA(p, q) class, where  $E(y_t | \mathcal{F}_{t-1}) = V(y_t | \mathcal{F}_{t-1}) = \mu_t$ .

Considering the link function, the linear predictor of this model is given by:

$$\log(\mu_t) = \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{j=1}^p \phi_j [\log(y_{t-j}) - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j [\log(y_{t-j}) - \log(\mu_{t-j})].$$

This can be written as the back-shift operator (B) (Briet, Amerasinghe, and Vounatsou 2013). In some cases, it is necessary to insert a threshold,  $c$ , to guarantee the existence of the link function. A possibility is to replace  $y_{t-j}$  by  $y_{t-j}^* = \max(y_{t-j}, c)$ ,  $c \in (0, 1)$ .

Given the conditioning of  $L(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Theta} | Y)$  to the first  $m$  observations of the series, we can suppose that the  $m$  first errors are zero. Moreover, the residuals associated with the moving average term can be constructed by using the Pearson residuals, residuals on the original scale, or residuals on the predictor scale (Benjamin *et al.* 2003; Rocha and Cribari-Neto 2008).

### 2.2. Negative binomial

We supposed that  $y_t | \mathcal{F}_{t-1}$  follows a Negative binomial distribution, i.e.  $\text{NB}(\mu_t, \sigma)$ . The conditional density is given by Equation 3:

$$p(y_t | \mathcal{F}_{t-1}) = \frac{\Gamma(y_t + \frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(y_t + 1)} \left( \frac{\sigma \mu_t}{1 + \sigma \mu_t} \right)^{y_t} \left( \frac{1}{1 + \sigma \mu_t} \right)^{\frac{1}{\sigma}}, \quad (3)$$

defined in  $\Omega = \{\mu_t, \sigma | \mu_t, \sigma > 0\}$  for  $y_t \in \mathbb{N}$ . There are several forms of the Negative binomial in the literature. Andrade *et al.* (2015), for example, considered the form where  $\nu = 1/\sigma$  is supposed to be known, belonging to the exponential family and resulting in a GARMA(p, q) submodel.

We relaxed the assumption of the GARMA(p, q), in which  $p(y_t | \mathcal{F}_{t-1})$  belongs to the exponential family, assuming that the dispersion parameter ( $\sigma$ ) is unknown. Therefore, the conditional mean is given by  $E(y_t | \mathcal{F}_{t-1}) = \mu_t$  and the conditional variance is equal to  $V(y_t | \mathcal{F}_{t-1}) = \mu_t(1 + \sigma \mu_t)$ .

Considering the logarithmic link function to ensure that  $\mu_t \subseteq \Omega$ , the linear predictor can be written as follows:

$$\log(\mu_t) = \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{j=1}^p \phi_j [\log(y_{t-j}) - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j [\log(y_{t-j}) - \log(\mu_{t-j})],$$

which is equal to the predictor of the Poisson, with the same restriction replacing  $y_{t-j}$  by  $y_{t-j}^* = \max(y_{t-j}, c)$ ,  $c \in (0, 1)$ . Given the conditioning of  $L(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \sigma | Y)$ , the assumption that the  $m$  first errors are zero following the conditions established by the Poisson model.

### 2.3. Poisson inverse Gaussian

4

Supposing that  $y_t | \mathcal{F}_{t-1}$  follows a Poisson inverse Gaussian,  $\text{PIG}(\mu_t, \sigma)$ , the conditional density is:

$$p(y_t | \mathcal{F}_{t-1}) = \left(\frac{2\alpha}{\pi}\right)^{\frac{1}{2}} \mu_t^{y_t} e^{\frac{1}{\sigma}} \left(K_{y_t - \frac{1}{2}}(\alpha)\right) \frac{1}{(\alpha\sigma)^{y_t} y_t!}, \quad (4)$$

where  $\Omega = \{\mu_t, \sigma | \mu_t, \sigma > 0\}$  for  $y_t \in \mathbb{N}$ . In equation 4,  $\alpha^2 = \frac{1}{\sigma^2} + \frac{2\mu_t}{\sigma}$  and  $K_\lambda(T) = \frac{1}{2} \int_0^\infty x^{(\lambda-1)} \exp\left\{-\frac{T}{2}(x+x^{-1})\right\} dx$ , where  $K(\cdot)$  is the modified Bessel function of the third kind. The conditional mean and variance are similar to the Negative binomial, that is,  $E(y_t | \mathcal{F}_{t-1}) = \mu_t$  and  $V(y_t | \mathcal{F}_{t-1}) = \mu_t(1 + \mu_t\sigma)$ . We consider the same linear predictor as shown in Poisson and the approximated likelihood is analogous to the Negative binomial model.

Due to the complexity of the likelihood functions of the Poisson, Negative binomial, and Poisson inverse Gaussian models, iterative methods can be used to estimate the parameters. In this paper, we consider a Bayesian analysis, justified by the inferential gain obtained in [Andrade et al. \(2015\)](#) and by the possibility of inserting prior knowledge.

### 3. Bayesian analysis

The Bayesian inference is based on the Bayes Theorem, where the posterior distribution is proportional to  $L(\boldsymbol{\theta} | Y)p(\boldsymbol{\theta})$ , being  $p(\boldsymbol{\theta})$  the joint prior distribution of  $\boldsymbol{\theta}$ . For convenience, consider  $\boldsymbol{\theta}$  as  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Theta})^\top$  in Poisson and  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \sigma)^\top$  in Negative binomial and Poisson inverse Gaussian models.

For the parameters that correspond to the effects of explanatory variables, we suppose that each component of  $\boldsymbol{\beta}$  is normally distributed, i.e.:

$$p(\beta_j) \propto \exp\left[-\frac{1}{2}\left(\frac{\beta_j - \mu_j}{\tau_j}\right)^2\right], \quad \beta_j \in (-\infty, \infty),$$

for  $j = \{1, \dots, r\}$ , being  $\mu_j$  and  $\tau_j$  the hyperparameters associated to  $\beta_j$ . A similar structure was adopted for  $\boldsymbol{\Phi}$  and  $\boldsymbol{\Theta}$ . It is:

$$p(\phi_k) \propto \exp\left[-\frac{1}{2}\left(\frac{\phi_k - \mu_{0k}}{\tau_{0k}}\right)^2\right], \quad \phi_k \in (-\infty, \infty),$$

$$p(\theta_l) \propto \exp\left[-\frac{1}{2}\left(\frac{\theta_l - \mu_{1l}}{\tau_{1l}}\right)^2\right], \quad \theta_l \in (-\infty, \infty),$$

where  $k = \{1, \dots, p\}$  and  $l = \{1, \dots, q\}$ . The hyperparameters were fixed at  $\mu_j = \mu_{0k} = \mu_{1l} = 0$  and  $\tau_j = \tau_{0k} = \tau_{1l} = 100$ , resulting in flat densities.

For the dispersion parameter of the Negative binomial and Poisson inverse Gaussian models, we considered a non-informative  $\text{Gamma}(s, a)$  prior with hyperparameters  $a = 1$  e  $s = 100^{-1}$ , that is given by:

$$p(\sigma) = \frac{\sigma^{a-1}}{s^a \Gamma(a)} \exp\left[-\frac{\sigma}{s}\right], \quad \sigma \in (0, \infty).$$

In Figure 1, we present the behavior of the prior densities of  $\beta_j$  (a) and  $\sigma$  (b). Analogously, the behavior of the densities of  $\phi_k$  and  $\theta_l$  is similar to that shown in Figure 1(a).

Considering the algebraic complexity of the joint posterior distributions, namely  $\pi(\boldsymbol{\theta} | Y)$ , the inference procedure can be performed by using Markov chain Monte Carlo (MCMC) methods, such as the Metropolis Hastings (MH) to sample from the joint posterior distribution, or using iterative algorithms for numerical optimization.



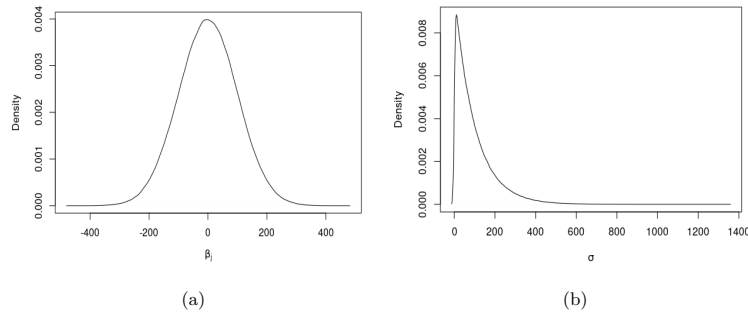


Figure 1: Prior densities of  $\beta_j$  and  $\sigma$ , respectively, that were used in the Bayesian analysis.

According to Korn, Korn, and Kroisandt (2010), the basic idea of the MCMC methods is to draw samples from a target distribution by simulating a Markov chain, in which the stationary distribution follows the target. The MH is a popular MCMC algorithm, proposed in Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) and generalized by Hastings (1970). It builds a chain started in an arbitrary state and has a transitional probability  $p_{i,j}$ , representing the movement probability of the state  $i$  to  $j$ , and satisfying the reversible equation. Defining  $q(i, j)$  as the transition kernel, the MH generates candidates from the transition and evaluates their acceptance probability (Korn *et al.* 2010). Because this process is dynamic, the chain will reach the stationary distribution over time. For properties and extensions of the MH algorithm, see Hastings (1970) and Korn *et al.* (2010).

### 3.1. Predictive density

In this Subsection, we derived the density of  $Y_{t+h}$  conditioned on all parameters and past observations, which is also called predictive density. Combining the joint posterior  $\pi(\boldsymbol{\theta} | Y)$  with the density of the future observation,  $y_{t+h}$ ,  $p(y_{t+h} | \boldsymbol{\theta}, \mathcal{F}_{t+h-1})$ , the predictive density is:

$$p(y_{t+h} | \mathcal{F}_{t+h-1}) = \int_{\boldsymbol{\theta} \in \Omega} p(y_{t+h} | \boldsymbol{\theta}, \mathcal{F}_{t+h-1}) \pi(\boldsymbol{\theta} | Y) d\boldsymbol{\theta},$$

which does not have a closed form. In this case, one strategy is to produce a Monte Carlo approximation of the predictive, drawing  $\mathcal{N}$  samples from  $\boldsymbol{\theta}^i$ ,  $i = \{1, \dots, \mathcal{N}\}$ , as follows:

$$p(y_{t+h} | \mathcal{F}_{t+h-1}) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} p(y_{t+h} | \boldsymbol{\theta}^i, \mathcal{F}_{t+h-1}).$$

This procedure was also performed by Andrade *et al.* (2015) and discussed in Gamerman and Lopes (2006); Krüger, Lerch, Thorarinsdottir, and Gneiting (2020). On this wise, the expected value of  $y_{t+h}$  is:

$$E(y_{t+h}) = \int_{y_{t+h} \in \Omega} y_{t+h} p(y_{t+h} | \mathcal{F}_{t+h-1}) dy_{t+h},$$

which can be approximated from  $\mu_{t+h}$ , drawing  $\mathcal{N}$  samples of  $\boldsymbol{\theta}^i$ ,  $i = \{1, \dots, \mathcal{N}\}$ , i.e.:

$$\hat{y}_{t+h} \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \mu_{t+h}(\boldsymbol{\theta}^i, \mathcal{F}_{t+h-1}).$$

## 4. A simulation study

In this Section, we implement a simulation study to analyze the performance of the Bayesian approach to POI-AR(1), NB-AR(1), and PIG-AR(1) models. Some parameter settings are taken into account, three sample sizes  $n = \{75; 125; 225\}$  and each model will be replicated  $w = 1,000$  times.

To simulate the  $w$  series, we used the `gamlss.dist` package (Stasinopoulos and Rigby 2020) with the inverse transform method, available in the R software (R Core Team 2021). Data was generated following its respective distributions and for the NB-AR(1) and PIG-AR(1) models, we fixed the dispersion parameter at  $\sigma = 0.25$ . We used the MH to estimate the parameters by using the package `MHadaptive` of Chivers (2015). The code was executed in Python, with the library `rpy2` (Gautier 2021). Configurations such as burn-in, thin, and total samples were determined by a pilot study.

The convergence was assessed in each replicated model via HW (Heidelberger and Welch 1983), G (Geweke 1992), and the Dependence Factor (I) (Raftery and Lewis 1992) criteria. To evaluate the convergence, we adopted an  $\alpha$  level of 0.05 in the HW diagnosis and a threshold equal to 0.10 (eps) in the half-width; we compared the  $|G|$  statistic with the  $Z_{1-\frac{\alpha}{2}}$  quantile and verified if the value of I tended to one.

In the simulation step, only the processes in which the convergence was obtained, according to the three criteria shown above, were considered. We evaluated the performance of the inference by using the Corrected Bias (CB) and Corrected Error (CE) in a similar way as Andrade *et al.* (2015). Those metrics were estimated by  $CE^2 = \frac{1}{w\tau^2} \sum_{i=1}^w (\hat{\theta}^i - \theta)^2$  e  $CB = \frac{1}{w} \sum_{i=1}^w \left| \frac{\theta - \hat{\theta}^i}{\theta} \right|$ , being  $\tau$  the standard deviation of  $\theta$  among the  $w$  replicates.

In Algorithm 1, we described our simulation procedure and some simulation results are available in Subsection 4.1. In the Appendix we presented the settings of the MH, including the burn-in period and convergence criteria. For convenience, we presented the mean of the convergence criteria and the acceptance probability.

---

### Algorithm 1 Steps of the simulation process

---

- 1: **Start**
  - 2: Set the model, the number of samples, burn-in and thin period.
  - 3: Set the parameters and  $n$  for generating the data.
  - 4: **for**  $w$  in 1 until  $w = 1,000$  **do**
  - 5:     Draw the series  $Y$  of length  $n$ .
  - 6:     Start the MH according to step (2).
  - 7:     Evaluate the convergence using the HW, G, and I criteria.
  - 8:     **if** the criteria indicate convergence **then**
  - 9:         Estimate the mean, mode, and standard deviation of the posterior distributions.
  - 10:         Store (9) and the convergence results.
  - 11:         Store the acceptance probability and the Monte Carlo error.
  - 12:          $w = w + 1$ .
  - 13:     **else**
  - 14:          $w = w$ .
  - 15:         Return to step 5.
  - 16:     **end if**
  - 17: **end for**
  - 18: Save the average of the  $w$  results of the posterior mean, mode, and standard deviation. Name these quantities as mean, mode, and SD, respectively.
  - 19: **End**
- 

### 4.1. Parameter settings and results

We simulated the series according to the following equation:

$$\begin{aligned} y_t | \mathcal{F}_{t-1} &\sim \text{POI} \{ \log(\mu_t) = \beta_1 + \phi_1 [\log(y_{t-1}^*) - \beta_1] \}. \\ y_t | \mathcal{F}_{t-1} &\sim \text{NB} \{ \log(\mu_t) = \beta_1 + \phi_1 [\log(y_{t-1}^*) - \beta_1], \sigma \}. \\ y_t | \mathcal{F}_{t-1} &\sim \text{PIG} \{ \log(\mu_t) = \beta_1 + \phi_1 [\log(y_{t-1}^*) - \beta_1], \sigma \}. \end{aligned} \quad (5)$$

where  $\beta = (\beta_1)^\top$  is the vector associated with the level and  $\mathbf{x}_t = (\mathbf{x}_1)$ , being  $\mathbf{x}_1 = (1, 1, \dots, 1)^\top$ . Table 1, shows the values used for the simulation of the artificial series.

Table 1: Parameter values used to generate the artificial time series.

Scenario	n	$\beta_1$	$\phi_1$	$\sigma$
I	{75, 125, 225}	1.00	0.10	0.25
II	{75, 125, 225}	1.00	0.40	0.25
III	{75, 125, 225}	1.00	0.70	0.25

In Table 2, the results of the first scenario are presented, where the parameter  $\phi_1 = 0.10$ . In general, there were reductions in the CB values when the sample size increased, suggesting good properties in the estimating process. This was also seen in the CE values, which tended to one.

The estimates of the dispersion parameter,  $\sigma$ , in NB-AR(1) and PIG-AR(1) models were improved as  $n$  increased. Our findings suggest a better performance of the modes estimates when compared to the averages to infer about the parameter  $\sigma$ , producing a good approximation when  $n = 225$ .

The other scenarios are included in the Appendix. When increasing  $\phi_1$  to 0.40, available in Table 7, the results of CB and CE metrics were similar to the first one, indicating improvement of the estimates with the increase in sample size. The estimates of the mode of the marginal distribution of  $\sigma$  were closer to the real value when compared to the average estimates.

Table 2: Main results of the first scenario for POI-AR(1), NB-AR(1), and PIG-AR(1) models based on the  $w$  replications, where  $\beta_1 = 1.00$ ,  $\phi_1 = 0.10$ , and  $\sigma = 0.25$ .

Model	n	Parameter	Mean	Mode	SD	CB	CE
POI-AR(1)	75	$\beta_1$	0.996	0.993	0.083	0.067	1.001
		$\phi_1$	0.113	0.109	0.076	0.601	1.015
	125	$\beta_1$	0.996	0.993	0.067	0.054	1.003
		$\phi_1$	0.105	0.103	0.059	0.471	1.004
	225	$\beta_1$	0.998	0.997	0.047	0.038	1.001
		$\phi_1$	0.102	0.101	0.043	0.346	1.001
NB-AR(1)	75	$\beta_1$	1.005	0.995	0.113	0.089	1.001
		$\phi_1$	0.118	0.116	0.077	0.626	1.027
		$\sigma$	0.309	0.271	0.121	0.403	1.113
	125	$\beta_1$	0.997	0.992	0.084	0.067	1.000
		$\phi_1$	0.109	0.108	0.059	0.467	1.011
		$\sigma$	0.289	0.261	0.088	0.293	1.064
	225	$\beta_1$	0.997	0.995	0.062	0.049	1.000
		$\phi_1$	0.104	0.103	0.045	0.362	1.003
		$\sigma$	0.266	0.255	0.066	0.216	1.029
PIG-AR(1)	75	$\beta_1$	1.018	1.004	0.112	0.090	1.012
		$\phi_1$	0.114	0.111	0.081	0.639	1.014
		$\sigma$	0.348	0.278	0.158	0.541	1.176
	125	$\beta_1$	1.005	0.998	0.087	0.070	1.001
		$\phi_1$	0.107	0.107	0.060	0.473	1.006
		$\sigma$	0.307	0.270	0.109	0.373	1.127
	225	$\beta_1$	1.006	1.002	0.064	0.051	1.003
		$\phi_1$	0.103	0.102	0.046	0.366	1.001
		$\sigma$	0.275	0.256	0.074	0.245	1.053

Similar results were found in the second and third scenarios, where the estimates of  $\phi_1$  improved as the sample size increased, indicating the necessity of large samples sizes, mainly in models with dispersion parameters. The results of these scenarios can be seen in Tables 7 and 8, respectively.

The importance of time series with large sample sizes was noted by Barreto-Souza (2017) when studying the PIG model with INAR(1) structure, reducing the bias and the standard errors as  $n$  increased. Regarding the scenarios studied in this paper, series in which  $n \ll 75$ , with the same linear predictor, are not considered interesting for modeling.

We verified a high computational cost with the MH algorithm in settings where  $\phi_1$  was greater than 0.80. In the situation where  $\phi_1 = 0.90$  and  $n = 225$  in the PIG-AR(1), the pilot model with thin = 20 and 80,000 total samples resulted in a low acceptance rate and a high values of I, suggesting the need to reparametrize or adopt another sampling technique.

## 5. Real data analysis

We considered the number of vehicle thefts in the region of *Campinas, São Paulo*, Brazil, from January 2010 to May 2022. It is a monthly time series, made available by the *Secretaria de Segurança Pública do Estado de São Paulo* (2021). For additional details about the database, visit the following website: <http://www.ssp.sp.gov.br/estatistica/pesquisa.aspx>. The series contains data from 38 localities near the city of *Campinas*, state of *São Paulo*, showed graphically in Figure 2:

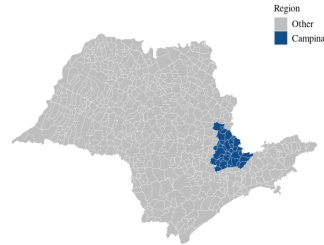


Figure 2: Geographic location of the cities included in the *Campinas* region. State of *São Paulo*, Brazil.

The samples from April and May 2022 were removed from the training stage, being used to perform an out-of-sample forecasting analysis. A brief summary of the data is available in Table 3. Note that the skewness and kurtosis measures indicate an asymmetric and leptokurtic behavior of the data.

Table 3: Descriptive statistics of the number of thefts reported between January 2010 and May 2022 in *Campinas*, Brazil.

Period	Mean	Median	SD	Skewness	Kurtosis
Jan/2010 - May/2022	691.380	700.000	147.949	-0.211	-0.516

We used a trend test, proposed by [Cox and Stuart \(1955\)](#), to evaluate the trend component. Graphically, this component can be verified by analyzing the behavior of the series over time, presented in Figure 3 (a). The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) of  $\Delta Y$  (series adjusted for trend) is shown in Figure 3 (b) and 3 (c).

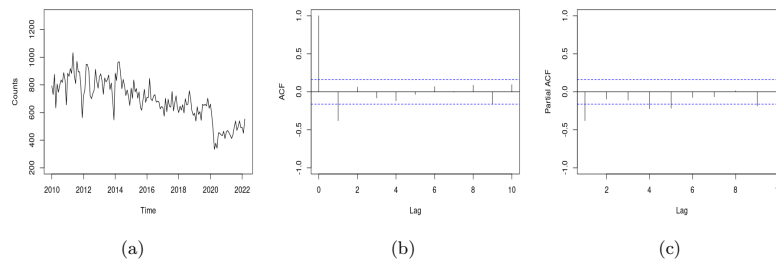


Figure 3: (a): Number of thefts reported in the region of *Campinas*; (b): Autocorrelation function of  $\Delta Y$ ; (c): Partial autocorrelation function of  $\Delta Y$ .

To model this series, it was considered a linear predictor with level and trend control and an AR(1) dependence structure, which is similar to Equation 5. However, in this case  $\beta = (\beta_1, \beta_2)^\top$  and  $\mathbf{x}_t = (\mathbf{x}_1, \mathbf{x}_2)^\top$ ,  $\mathbf{x}_1 = (1, 1, \dots, 1)^\top$  and  $\mathbf{x}_2 = (1, 2, \dots, n)^\top$ . The MH algorithm was used for sampling from the joint posterior, setting the burn-in period equal to 2,000, 10,000 total samples and keeping the twentieth sample value as the thinning period. The estimation results are in Table 4 and the graph of the posterior marginal density of each parameter is available in Figures 4, 5, and 6. In relation to the convergence to the stationary distribution, all the processes met the established criteria.

We compared the Bayesian estimates with the maximum likelihood (ML) ones. The likelihood functions were numerically optimized using the L-BFGS-B algorithm (Byrd, Lu, Nocedal, and Zhu 1995) and the standard errors were computed based on the inverse of the Hessian matrix. When analyzing the ML results, available in Table 12, we can verify the proximity of the estimates between these approaches, especially for the parameter  $\beta_2$  of the POI-AR(1) and NB-AR(1) models. The greatest distances between the Bayesian and maximum likelihood estimates were observed in  $\beta_1$  and  $\phi_1$  of the PIG-AR(1) model.

Table 4: Estimation results for POI-AR(1), NB-AR(1), and PIG-AR(1) models for the number of thefts in *Campinas, São Paulo*, Brazil.

Model	Parameter	Mean	Mode	SD	HPD (95%)	
					$L_l$	$L_u$
POI-AR(1)	$\beta_1$	6.8346	6.8342	0.0128	6.8107	6.8604
	$\beta_2$	-0.0042	-0.0041	0.0002	-0.0045	-0.0038
	$\phi_1$	0.5246	0.5227	0.0241	0.4776	0.5716
NB-AR(1)	$\beta_1$	6.8534	6.8546	0.0465	6.7604	6.9437
	$\beta_2$	-0.0044	-0.0044	0.0006	-0.0055	-0.0033
	$\phi_1$	0.5663	0.5560	0.0706	0.4238	0.7010
	$\sigma$	0.0116	0.0114	0.0016	0.0087	0.0149
PIG-AR(1)	$\beta_1$	6.8517	6.8503	0.0486	6.7525	6.9456
	$\beta_2$	-0.0044	-0.0044	0.0006	-0.0055	-0.0033
	$\phi_1$	0.5816	0.5817	0.0723	0.4380	0.7192
	$\sigma$	0.0119	0.0115	0.0016	0.0089	0.0151

Considering the results shown in Table 4, we can observe the significant trend effect in the three models, indicating a reduction in the number of thefts during the analyzed period. Furthermore, the credible intervals of  $\beta_1$  and  $\phi_1$  were wider in NB-AR(1) and PIG-AR(1) models.

In PIG-AR(1), the dispersion parameter was estimated in 0.012,  $CI_\sigma = \{0.089; 0.015\}$ , and a similar result occurred with NB-AR(1). It indicates that the variability of the series is greater than the mean and the phenomenon has slightly heavier tails. According to these models, the conditional variance is given by  $V(y_t | \mathcal{F}_{t-1}) \approx \mu_t(1 + 0.012\mu_t)$ , for all  $t$ .

The deviance information criterion (DIC) and the conditional predictive ordinate (CPO) results are available in Table 5, suggesting a preference for mixed distributions when compared to Poisson, since they presented lower DIC and higher CPO values.

Table 5: Deviance information criterion and the conditional predictive ordinate for POI-AR(1), NB-AR(1), and PIG-AR(1) models.

Model	POI-AR(1)	NB-AR(1)	PIG-AR(1)
DIC	6232.812	<b>4328.217</b>	4332.140
CPO	-1245.136	<b>-842.182</b>	-843.139

We proceeded with the analysis estimating the randomized quantile residuals proposed by Dunn and Smyth (1996), defined as  $r_t = \Phi^{-1}(F_{y_t}(y_t | \mathcal{F}_{t-1}))$ , where  $\Phi^{-1}$  is the inverse cumulative distribution function of a standard normal and  $F_{y_t}$  is the fitted conditional distribution that was used for modeling. According to the Box and Pierce (1970) test, based on 20 lags, there is no evidence of any correlation within the residuals (p-values equal to 0.419, 0.127, and 0.128 for POI-AR(1), NB-AR(1), and PIG-AR(1), respectively). The Kolmogorov-Smirnov test did not reject the null hypothesis of normality of the residuals, returning p-values equal to 0.123, 0.414, and 0.421.

Predictions for April and May 2022 are shown in Table 6. We estimated the Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) for these predictions. The values

of MAPE and MAE were  $\text{MAPE} = \{14.55; 14.70; 14.68\}$  and  $\text{MAE} = \{60.60; 61.67; 61.32\}$  for POI-AR(1), NB-AR(1), and PIG-AR(1), respectively. Overall, mixed models obtained good results according to the information criteria, but the MAPE and MAE metrics were similar to Poisson's. In addition to this result, the predictions made with PIG-AR(1) showed a lower error when compared with the predictions of NB-AR(1).

Table 6: Predictions with POI-AR(1), NB-AR(1), and PIG-AR(1) models for the number of thefts in *Campinas, São Paulo, Brazil*.

Model	Month	Real value	Predicted	SD	HPD (95%)	
					$L_l$	$L_u$
POI-AR(1)	April	412	526.916	3.880	519.545	534.660
	May	519	512.711	3.692	505.345	519.731
NB-AR(1)	April	412	524.738	11.201	502.088	545.575
	May	519	508.397	10.452	488.904	529.391
PIG-AR(1)	April	412	526.728	11.106	505.626	549.406
	May	519	511.083	10.364	491.457	532.176

The predictions for April were overestimated, with values consistently above the real counts, and the credible intervals did not include those values. The prediction for May, based on the PIG-AR(1) model, was relatively close to the real value and the April prediction was overestimated. The same happened with the NB-AR(1) model.

## 6. Final remarks

In this paper we investigated models of count time series using Poisson, Negative binomial, and Poisson inverse Gaussian distributions, the last two being alternatives for overdispersed count data. We extended the Bayesian analysis presented in the literature to the Poisson inverse Gaussian and relaxed the assumption of known dispersion parameter in the GARMA(p, q) class for the Negative binomial distribution.

Through a simulation study, we verified the advantages of the Bayesian inference in terms of implementation, estimation, and flexibility of the models, as well as the possibility of modeling the dispersion parameter in the mixed distributions. The Bayesian perspective works well for the given scenarios. The analysis with real data showed that the Poisson inverse Gaussian distribution is an alternative for modeling count data, providing good predictions.

Our possible points for future research include constructing the Probability Integral Transform (PIT) of [Czado, Gneiting, and Held \(2009\)](#) and studying higher AR(p) orders in the Poisson inverse Gaussian model using the Bayesian inference, and considering the Hamiltonian Monte Carlo (HMC) algorithm to draw samples from the posterior distribution.

## Declaration of conflicting interests

The authors declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

## Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

## References

- Andrade BS, Andrade MG, Ehlers RS (2015). “Bayesian GARMA Models for Count Data.” *Communications in Statistics: Case Studies, Data Analysis and Applications*, **1**(4), 192–205. doi:10.1080/23737484.2016.1190307.
- Barreto-Souza W (2017). “Mixed Poisson INAR(1) Processes.” *Statistical Papers*, **60**, 2119–2139. doi:10.1007/s00362-017-0912-x.
- Benjamin A, Rigby A, Stasinopoulos M (2003). “Generalized Autoregressive Moving Average Models.” *Journal of the American Statistical Association*, **98**(461), 214–223. doi:10.1198/016214503388619238.
- Box G, Jenkins G (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day. ISBN 0816211043.
- Box GE, Pierce DA (1970). “Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models.” *Journal of the American Statistical Association*, **65**(332), 1509–1526. doi:https://doi.org/10.2307/2284333.
- Briet OJT, Amerasinghe PH, Vounatsou P (2013). “Generalized Seasonal Autoregressive Integrated Moving Average Models for Count Data with Application to Malaria Time Series with Low Case Numbers.” *Plos One*, **8**(6), 1–9. doi:https://doi.org/10.1371/journal.pone.0065761.
- Byrd RH, Lu P, Nocedal J, Zhu C (1995). “A Limited Memory Algorithm for Bound Constrained Optimization.” *Journal on Scientific Computing*, **16**(5), 1190–1208. doi:10.1137/0916069.
- Chivers C (2015). *MHadaptive: General Markov Chain Monte Carlo for Bayesian Inference Using Adaptive Metropolis Hastings Sampling*. R package version 1.1-8, URL <https://rdrr.io/cran/MHadaptive/>.
- Cox DR, Gudmundsson G, Lindgren G, Bondesson L, Harsaae E, Laake P, Juselius K, Lauritzen SL (1981). “Statistical Analysis of Time Series: Recent Developments.” *Scandinavian Journal of Statistics*, **8**(2), 93–115.
- Cox DR, Stuart A (1955). “Some Quick Sign Tests for Trend in Location and Dispersion.” *Biometrika*, **42**(1/2), 80–95. doi:10.1093/biomet/42.1-2.80.
- Czado C, Gneiting T, Held L (2009). “Predictive Model Assessment for Count Data.” *Biometrics*, **65**(4), 1254–1261. doi:10.1111/j.1541-0420.2009.01191.x.
- Davis RA, Fokianos K, Holan SH, Joe H, Livsey J, Lund R, Pipiras V, Ravishanker N (2021). “Count Time Series: A Methodological Review.” *Journal of the American Statistical Association*, **116**(535), 1533–154. doi:10.1080/01621459.2021.1904957.
- Dean C, Lawless J, Willmot G (1989). “A mixed Poisson Inverse Gaussian Regression Model.” *Canadian Journal of Statistics*, **17**(2), 171–181. doi:https://doi.org/10.2307/3314846.
- Dunn PK, Smyth GK (1996). “Randomized Quantile Residuals.” *Journal of Computational and Graphical Statistics*, **5**(3), 236–244. doi:https://doi.org/10.2307/1390802.
- Dunsmuir W, Scott D (2015). “The Glarma Package for Observation-Driven Time Series Regression of Counts.” *Journal of Statistical Software*, **67**(7), 1–36. doi:10.18637/jss.v067.i07.
- Gamerman D, Lopes HF (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press. ISBN 1584885874.
- Gautier L (2021). *Rpy2: A Simple and Efficient Access to R from Python*. URL <https://rpy2.github.io/doc/latest/html/introduction.html>.



- Geweke JF (1992). “Evaluating the Accuracy of Sampling Based Approaches to the Calculation of Posterior Moments.” In J Bernardo, J Berger, A Dawid, A Smith (eds.), *Bayesian Statistics 4*, pp. 625–631. Oxford University Press.
- Gonçalves JN, Barreto-Souza W (2020). “Flexible Regression Models for Counts with High Inflation of Zeros.” *Metron*, **78**(1), 71–95. doi:10.1007/s40300-020-00163-9.
- Hastings WK (1970). “Monte Carlo Sampling Methods Using Markov Chains and their Applications.” *Biometrika*, **57**(1), 97–109. doi:https://doi.org/10.2307/2334940.
- Heidelberger P, Welch P (1983). “Simulation Run Length Control in the Presence of an Initial Transient.” *Operations Research*, **31**(6), 1109–1144. doi:10.1287/opre.31.6.1109.
- Khandelwal I, Adhikari R, Verm G (2015). “Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition.” *Procedia Computer Science*, **48**, 173–179. doi:10.1016/j.procs.2015.04.167.
- Kong J, Lund R (2022). “Seasonal Count Time Series.” *Journal of Time Series Analysis*, pp. 1–32. doi:10.1111/jtsa.12651.
- Korn R, Korn E, Kroisandt G (2010). *Monte Carlo Methods and Models in Finance and Insurance*. CRC press. ISBN 1420076183.
- Krüger F, Lerch S, Thorarinsdottir T, Gneiting T (2020). “Predictive Inference Based on Markov Chain Monte Carlo Output.” *International Statistical Review*, **89**(2), 274–301. doi:https://doi.org/10.1111/insr.12405.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953). “Equation of State Calculations by Fast Computing Machines.” *The Journal of Chemical Physics*, **21**(6), 1087–1092. doi:10.1063/1.1699114?ver=pdfcov.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Raftery AE, Lewis SM (1992). “Comment: One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo.” *Statistical Science*, **7**(4), 493–497.
- Rocha AV, Cribari-Neto F (2008). “Beta Autoregressive Moving Average Models.” *Test*, **18**(3), 529–545. doi:10.1007/s11749-008-0112-z.
- Sales LOF, Alencar AP, Ho LL (2022). “The BerG Generalized Autoregressive Moving Average Model for Count Time Series.” *Computers & Industrial Engineering*, **168**. doi:10.1016/j.cie.2022.108104.
- Secretaria de Segurança Pública do Estado de São Paulo (2021). *Estatísticas*. Secretaria de Segurança Pública do Estado de São Paulo. URL <http://www.ssp.sp.gov.br/>.
- Silva RR (2020). “Generalized Autoregressive Neural Network Models.” *Arxiv*. doi:10.48550/arXiv.2002.05676.
- Stasinopoulos M, Rigby R (2020). *Gamlss.dist: Distributions for Generalized Additive Models for Location Scale and Shape*. R package version 5.1-6, URL <https://CRAN.R-project.org/package=gamlss.dist>.
- Willmot GE (1987). “The Poisson Inverse Gaussian Distribution as an Alternative to the Negative Binomial.” *Scandinavian Actuarial Journal*, **1987**(3-4), 113–127. doi:10.1080/03461238.1987.10413823.

## 7. Appendix

### 7.1. Additional simulation results

Table 7: Main results of the second scenario for POI-AR(1), NB-AR(1), and PIG-AR(1) models based on the  $w$  replications, where  $\beta_1 = 1.00$ ,  $\phi_1 = 0.40$ , and  $\sigma = 0.25$ .

Model	n	Parameter	Mean	Mode	SD	CB	CE
POI-AR(1)	75	$\beta_1$	0.978	0.974	0.137	0.108	1.012
		$\phi_1$	0.376	0.370	0.080	0.166	1.044
	125	$\beta_1$	0.987	0.984	0.101	0.078	1.008
		$\phi_1$	0.395	0.392	0.065	0.131	1.002
	225	$\beta_1$	0.993	0.992	0.076	0.061	1.004
		$\phi_1$	0.400	0.399	0.051	0.102	1.000
NB-AR(1)	75	$\beta_1$	0.990	0.959	0.185	0.149	1.001
		$\phi_1$	0.389	0.385	0.084	0.171	1.007
		$\sigma$	0.343	0.292	0.140	0.503	1.202
	125	$\beta_1$	0.997	0.982	0.141	0.110	1.000
		$\phi_1$	0.402	0.399	0.066	0.132	1.000
		$\sigma$	0.293	0.265	0.103	0.339	1.083
	225	$\beta_1$	0.996	0.988	0.102	0.082	1.000
		$\phi_1$	0.401	0.400	0.050	0.101	1.000
		$\sigma$	0.264	0.250	0.077	0.245	1.016
PIG-AR(1)	75	$\beta_1$	1.042	1.000	0.172	0.141	1.029
		$\phi_1$	0.390	0.384	0.086	0.174	1.007
		$\sigma$	0.400	0.307	0.170	0.680	1.336
	125	$\beta_1$	1.005	0.985	0.140	0.112	1.000
		$\phi_1$	0.400	0.397	0.068	0.135	1.000
		$\sigma$	0.319	0.271	0.128	0.435	1.136
	225	$\beta_1$	1.006	0.997	0.105	0.083	1.001
		$\phi_1$	0.401	0.400	0.049	0.098	1.000
		$\sigma$	0.284	0.260	0.087	0.292	1.072

Table 8: Main results of the third scenario for POI-AR(1), NB-AR(1), and PIG-AR(1) models based on the  $w$  replications, where  $\beta_1 = 1.00$ ,  $\phi_1 = 0.70$ , and  $\sigma = 0.25$ .

Model	n	Parameter	Mean	Mode	SD	CB	CE
POI-AR(1)	75	$\beta_1$	0.970	0.965	0.272	0.208	1.006
		$\phi_1$	0.610	0.602	0.067	0.133	1.682
	125	$\beta_1$	0.967	0.968	0.218	0.171	1.011
		$\phi_1$	0.671	0.666	0.057	0.070	1.124
	225	$\beta_1$	0.963	0.965	0.173	0.140	1.022
		$\phi_1$	0.697	0.695	0.050	0.058	1.001
NB-AR(1)	75	$\beta_1$	1.088	0.979	0.381	0.312	1.026
		$\phi_1$	0.627	0.617	0.074	0.118	1.402
		$\sigma$	0.330	0.259	0.162	0.500	1.114
	125	$\beta_1$	1.040	0.960	0.321	0.256	1.007
		$\phi_1$	0.675	0.669	0.064	0.077	1.073
		$\sigma$	0.308	0.265	0.123	0.404	1.105
	225	$\beta_1$	0.986	0.940	0.264	0.207	1.001
		$\phi_1$	0.697	0.694	0.055	0.063	1.001
		$\sigma$	0.289	0.265	0.099	0.319	1.075
PIG-AR(1)	75	$\beta_1$	1.163	1.016	0.366	0.322	1.094
		$\phi_1$	0.618	0.607	0.072	0.125	1.508
		$\sigma$	0.394	0.269	0.210	0.700	1.212
	125	$\beta_1$	1.073	0.976	0.316	0.260	1.026
		$\phi_1$	0.660	0.653	0.061	0.081	1.197
		$\sigma$	0.363	0.286	0.167	0.576	1.207
	225	$\beta_1$	1.027	0.972	0.256	0.208	1.005
		$\phi_1$	0.700	0.696	0.053	0.061	1.000
		$\sigma$	0.306	0.267	0.114	0.387	1.115

## 7.2. Settings of the simulation procedure

Table 9: MH settings and convergence analysis in the first scenario based on the  $w$  replications. Where  $\overline{AC}$  is the mean of the acceptance rate,  $\overline{G}$  is the mean of the G statistic,  $\overline{I}$  is the mean of I, and  $\overline{HW}$  is the mean of the p-values of the HW test.

Model	n	Samples	Burn-in	Thin	$\overline{AC}$	$\theta$	$\overline{I}$	$\overline{G}$	$\overline{HW}$
POI-AR(1)	75	5,000	1,000	7	0.54	$\beta_1$	1.16	0.02	0.53
						$\phi_1$	1.28	0.01	0.51
	125	5,000	1,000	7	0.54	$\beta_1$	1.06	0.03	0.52
						$\phi_1$	1.12	0.01	0.53
	225	5,000	1,000	7	0.51	$\beta_1$	1.26	0.02	0.52
						$\phi_1$	1.25	0.01	0.53
NB-AR(1)	75	5,000	1,000	7	0.43	$\beta_1$	1.29	0.01	0.52
						$\phi_1$	1.32	0.01	0.52
	125	5,000	1,000	8	0.44	$\sigma$	1.17	0.02	0.52
						$\beta_1$	1.24	0.01	0.53
	225	5,000	1,000	7	0.45	$\phi_1$	1.26	0.01	0.52
						$\sigma$	1.14	0.06	0.51
PIG-AR(1)	75	5,000	1,000	8	0.42	$\beta_1$	1.29	0.02	0.53
						$\phi_1$	1.30	0.00	0.51
	125	5,000	1,000	8	0.43	$\sigma$	1.22	0.01	0.51
						$\beta_1$	1.24	0.01	0.52
	225	5,000	1,000	8	0.44	$\phi_1$	1.27	0.00	0.52
						$\sigma$	1.10	0.07	0.53
225	5,000	1,000	8	0.44	$\beta_1$	1.24	0.03	0.51	
					$\phi_1$	1.26	0.04	0.51	
225	5,000	1,000	8	0.44	$\sigma$	1.11	0.09	0.50	
					$\beta_1$	1.23	0.01	0.50	
225	5,000	1,000	8	0.44	$\phi_1$	1.25	0.01	0.51	
					$\sigma$	1.13	0.05	0.49	

Table 10: MH settings and convergence analysis in the second scenario based on the  $w$  replications. Where  $\overline{AC}$  is the mean of the acceptance rate,  $\overline{G}$  is the mean of the G statistic,  $\overline{I}$  is the mean of I, and  $\overline{HW}$  is the mean of the p-values of the HW test.

Model	n	Samples	Burn-in	Thin	$\overline{AC}$	$\theta$	$\overline{I}$	$\overline{G}$	$\overline{HW}$
POI-AR(1)	75	5,000	1,000	7	0.55	$\beta_1$	1.29	0.01	0.52
						$\phi_1$	1.23	0.01	0.52
	125	5,000	1,000	7	0.55	$\beta_1$	1.28	0.01	0.51
						$\phi_1$	1.23	0.05	0.51
	225	5,000	1,000	8	0.55	$\beta_1$	1.20	0.03	0.52
						$\phi_1$	1.18	0.03	0.51
NB-AR(1)	75	5,000	1,000	8	0.42	$\beta_1$	1.25	0.01	0.52
						$\phi_1$	1.26	0.00	0.53
						$\sigma$	1.12	0.01	0.52
	125	5,000	1,000	8	0.44	$\beta_1$	1.24	0.01	0.52
						$\phi_1$	1.25	0.02	0.52
						$\sigma$	1.13	0.06	0.51
225	5,000	1,000	8	0.44	$\beta_1$	1.21	0.04	0.49	
					$\phi_1$	1.22	0.00	0.51	
					$\sigma$	1.13	0.12	0.50	
PIG-AR(1)	75	5,000	1,000	7	0.40	$\beta_1$	1.29	0.07	0.51
						$\phi_1$	1.32	0.05	0.53
						$\sigma$	1.15	0.06	0.52
	125	5,000	1,000	8	0.43	$\beta_1$	1.24	0.00	0.52
						$\phi_1$	1.26	0.03	0.52
						$\sigma$	1.11	0.08	0.51
225	5,000	1,000	7	0.44	$\beta_1$	1.28	0.01	0.50	
					$\phi_1$	1.30	0.03	0.51	
					$\sigma$	1.18	0.04	0.51	

Table 11: MH settings and convergence analysis in the third scenario based on the  $w$  replications. Where  $\overline{AC}$  is the mean of the acceptance rate,  $\overline{G}$  is the mean of the G statistic,  $\overline{I}$  is the mean of I, and  $\overline{HW}$  is the mean of the p-values of the HW test.

Model	n	Samples	Burn-in	Thin	$\overline{AC}$	$\theta$	$\overline{I}$	$\overline{G}$	$\overline{HW}$
POI-AR(1)	75	7,000	1,000	12	0.51	$\beta_1$	1.29	0.04	0.51
						$\phi_1$	1.30	0.01	0.52
	125	5,000	1,000	10	0.51	$\beta_1$	1.24	0.03	0.51
						$\phi_1$	1.30	0.00	0.50
	225	5,000	1,000	10	0.52	$\beta_1$	1.25	0.01	0.52
						$\phi_1$	1.20	0.02	0.51
NB-AR(1)	75	7,000	1,000	7	0.37	$\beta_1$	1.30	0.15	0.52
						$\phi_1$	1.34	0.04	0.51
						$\sigma$	1.55	0.03	0.49
	125	7,000	1,000	8	0.39	$\beta_1$	1.01	0.06	0.51
						$\phi_1$	1.22	0.01	0.49
						$\sigma$	1.23	0.01	0.49
225	10,000	1,000	8	0.41	$\beta_1$	1.10	0.02	0.48	
					$\phi_1$	1.22	0.02	0.50	
					$\sigma$	1.23	0.01	0.49	
PIG-AR(1)	75	11,000	1,000	11	0.35	$\beta_1$	1.32	0.17	0.53
						$\phi_1$	1.25	0.08	0.50
						$\sigma$	1.23	0.13	0.52
	125	5,000	1,000	12	0.39	$\beta_1$	1.22	0.15	0.52
						$\phi_1$	1.23	0.02	0.52
						$\sigma$	1.14	0.13	0.51
225	5,000	1,000	12	0.40	$\beta_1$	1.26	0.04	0.49	
					$\phi_1$	1.21	0.02	0.49	
					$\sigma$	1.29	0.01	0.49	

## 7.3. Behavior of the marginal densities

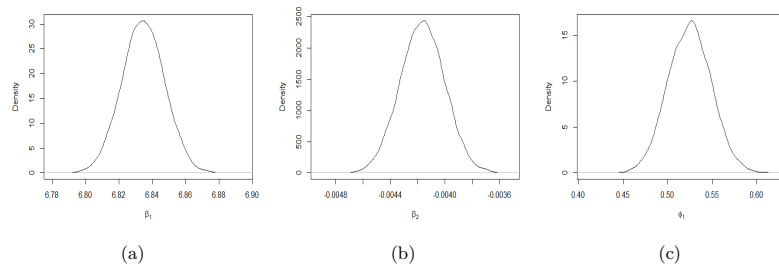


Figure 4: Posterior marginal densities of the POI-AR(1) parameters.

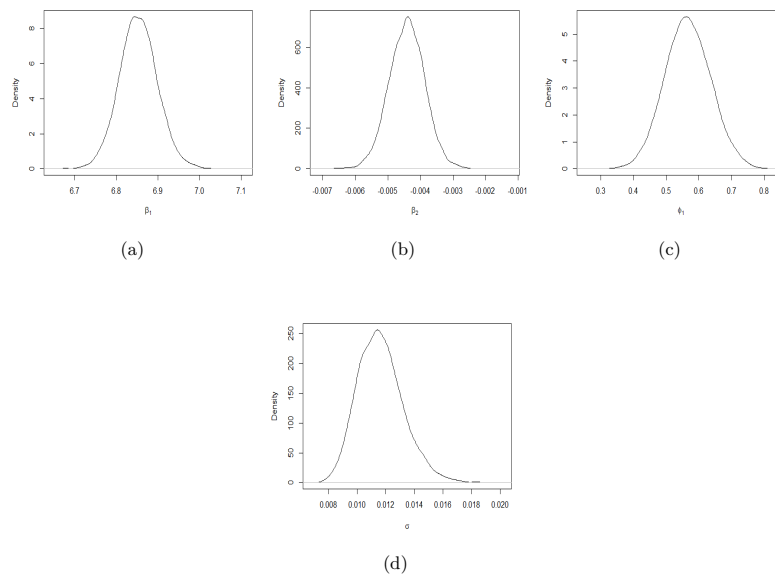


Figure 5: Posterior marginal densities of the NB-AR(1) parameters.

20

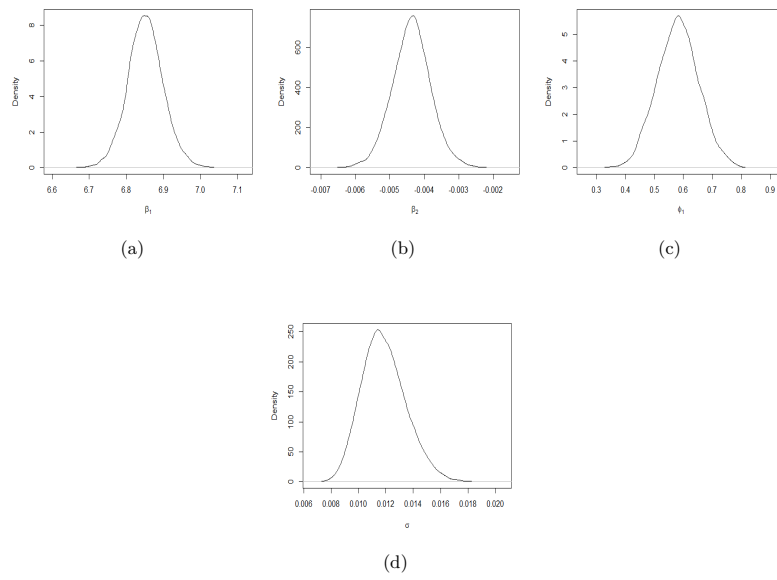


Figure 6: Posterior marginal densities of the PIG-AR(1) parameters.



## 7.4. Comparison with the maximum likelihood estimates

Table 12: Maximum likelihood estimates of POI-AR(1), NB-AR(1), and PIG-AR(1) models.

Model	Parameter	Estimated	Standard error
POI-AR(1)	$\beta_1$	6.8413	0.0128
	$\beta_2$	-0.0042	0.0002
	$\phi_1$	0.5235	0.0238
NB-AR(1)	$\beta_1$	6.8578	0.0420
	$\beta_2$	-0.0044	0.0005
	$\phi_1$	0.5443	0.0673
	$\sigma$	0.0110	0.0014
PIG-AR(1)	$\beta_1$	6.9841	0.1455
	$\beta_2$	-0.0058	0.0015
	$\phi_1$	0.6621	0.1460
	$\sigma$	0.0123	0.0021

**Affiliation:**

Luiz Otávio de Oliveira Pala  
 Department of Exact Sciences  
 Federal University of Lavras  
 Lavras, Minas Gerais, Brazil  
 E-mail: [luizotavio.oliveira@gmail.com](mailto:luizotavio.oliveira@gmail.com)  
 URL: <http://lattes.cnpq.br/5770791208849986>

**ARTIGO 2 - *Count time series with excess zeros: a Bayesian perspective using zero-adjusted distributions***

Redigido conforme as normas da revista *Semina: exact and technological sciences* (versão publicada).

ORIGINAL ARTICLE  
DOI: 10.5433/1679-0375.2022v43n2p147

## *Count time series with excess zeros: A Bayesian perspective using zero-adjusted distributions*

### *Séries temporais de contagem com excesso de zeros: Uma perspectiva Bayesiana utilizando distribuições zero-ajustadas*

Luiz Otávio de Oliveira Pala<sup>1</sup>; Marcela de Marillac Carvalho<sup>2</sup>;  
Thelma Sáfiadi<sup>3</sup>

#### Abstract

Models for count data which are temporally correlated have been studied using many conditional distributions, such as the Poisson distribution, and the insertion of different dependence structures. Nonetheless, excess of zeros and over dispersion may be observed during the counting process and need to be considered when modelling and choosing a conditional distribution. In this paper, we propose models for counting time series using zero-adjusted distributions by inserting a dependence structure following the ARMA(p, q) process on a Bayesian framework. We perform a simulation study using the proposed Bayesian analysis and analyse the monthly time series of the number of deaths due to dengue haemorrhagic fever (ICD-A91) in Brazil.

**Keywords:** ARMA(p, q) process; count data; metropolis-hastings.

#### Resumo

Modelos para dados de contagem temporalmente correlacionados têm sido estudados utilizando diversas distribuições condicionais, como a Poisson, e com a inserção de diferentes estruturas de dependência. No entanto, os fenômenos de contagem podem apresentar características como excesso de zeros e alta dispersão, que devem ser levadas em consideração durante a modelagem e escolha de uma distribuição condicional. Este trabalho tem como objetivo estudar modelos para séries de contagem utilizando três distribuições condicionais zero-ajustadas com estruturas de dependência na forma ARMA(p, q), em uma perspectiva via inferência Bayesiana. De forma geral, foi realizado um breve estudo de simulação a partir da análise Bayesiana proposta e a série temporal do número de óbitos em decorrência de febre hemorrágica causada pelo vírus da dengue (CID-A91) no Brasil foi analisada.

**Palavras-chave:** processo ARMA(p, q); dados de contagem; metropolis-hastings.

<sup>1</sup> PhD student in Statistics and Agricultural Experimentation, UFLA, Lavras, MG, Brazil, E-mail: luizotavio.oliveira@gmail.com

<sup>2</sup> Dr., Axxin Data Intelligence, Varginha, MG, Brasil, E-mail: marcela-carvalho\_@hotmail.com

<sup>3</sup> Prof. Dr., Department of Statistics, UFLA, Lavras, MG, Brazil, E-mail: safadi@ufla.br

## Introduction

Count data are used and analysed in areas such as economy, epidemiology, public health and many other applied areas (HASHIM; HASHIM; SHIKER, 2021; ZUO *et al.*, 2021). The modelling of these types of data are commonly based on Generalized Linear Models (GLMs) introduced by Nelder and Wedderburn (1972), with a certain predominance of Poisson regression (GONÇALVES; BARRETO-SOUZA, 2020).

However, the assumption of equidispersion arising from the Poisson distribution is rarely satisfied (HILBE, 2014). This is because over dispersion is commonly observed in count data (BARRETO-SOUZA, 2017), resulting from extra population heterogeneity, the omission of important explanatory variables, the presence of atypical observations (PAYNE *et al.*, 2017) and an excess of zeros (BARRETO-SOUZA, 2017); high dispersion needs to be handled during modelling because it can lead to inferential problems (PAYNE *et al.*, 2017).

Regression models based on mixed Poisson distributions have become attractive for treating over dispersion by inserting a latent random effect into the mean of the Poisson distribution (BARRETO-SOUZA, 2017). Models with Negative binomial and Poisson inverse Gaussian distributions are examples of this approach (BARRETO-SOUZA, 2017; HILBE, 2014).

Excess zeros can still be observed in the applied areas (FENG, 2021), such as in disease monitoring, use of health services and number of daily deaths from certain causes (HASHIM; HASHIM; SHIKER, 2021; TAWIAH; IDDRISU; ASOSEGA, 2021; YANG; ZAMBA; CAVANAUGH, 2013), a fact that should not be ignored, since it can negatively impact the inference, for example, generating spurious relationships (ALQAWBA; DIAWARA; CHAGANTY, 2019; YANG; ZAMBA; CAVANAUGH, 2013).

For modelling count time series, Benjamin, Rigby and Stasinopoulos (2003) proposed the Generalized Autoregressive Moving Average - GARMA(p, q) - class, based on the GLM theory, inserting a dependence structure following the ARMA(p, q) process, which has been used in models with Poisson, Binomial and Negative binomial distributions.

The literature on counting time series considering excess zeros is still sparse (YANG; ZAMBA; CAVANAUGH, 2013) but is a topic of recent interest in statistics (ALQAWBA; DIAWARA; CHAGANTY, 2019).

Yang, Zamba and Cavanaugh (2013), Alqawba, Diawara and Chaganty (2019), Ghahramani and White (2020), Tawiah, Iddrisu and Asosega (2021) are examples of studies that have modelled count time series with many zeros.

In the study by Yang, Zamba and Cavanaugh (2013), the zero-inflated Poisson (ZIP) distribution was used by incorporating lags of the response variable in the linear predictor to handle the temporal correlation in syphilis data. The Poisson, Negative binomial and Conway-Maxwell-Poisson distributions were adopted by Alqawba, Diawara and Chaganty (2019), who supposed that the errors of the regression follow a stationary ARMA(p, q) process.

Ghahramani and White (2020) studied the ZIP and zero-inflated Negative binomial (ZINB) models and developed a distribution-free approach when modelling the time series of syphilis cases. Tawiah, Iddrisu and Asosega (2021) analysed the COVID-19 mortality series using the ZIP and ZINB distributions with the ARMA(1, 0) dependence structure.

In this paper, we aim to study models for counting time series with excess zeros using Poisson, Negative binomial and Poisson inverse Gaussian zero-adjusted distributions, which are alternatives to zero-inflated distributions. We adopt the Bayesian approach, as in the study by Andrade, Andrade and Ehlers (2015), when modelling series using the GARMA(p, q) class.

Initially, we present the concept of zero-adjusted distributions and define models with the insertion of the ARMA(p, q) dependence structure. Then, we perform a simulation based on the proposed Bayesian analysis and analyse the time series of mortality due to dengue haemorrhagic fever (ICD-A91), according to the 10th revision of the International Classification of Diseases (ICD-10) (WORLD HEALTH ORGANIZATION, 2004). We also present some final remarks and suggestions for future work.

## Background on zero-adjusted and zero-inflated distributions

When dealing with data with a high proportion of zeros, zero-inflated and zero-altered distributions, also referred to as hurdle and adjusted distributions, are used for modelling and inference (FENG, 2021; HASHIM; HASHIM; SHIKER, 2021). According to Feng (2021), there is a conceptual and parameter interpretation difference when considering zero-inflated or hurdle regression.

In zero-inflated models, which were proposed by Lambert (1992), it is assumed that the zeros come from two processes. The first process is composed of situations with structural zeros, and the second is composed of random zeros, which are modelled by a count distribution such as the Poisson distribution (FENG, 2021; HASHIM; HASHIM; SHIKER, 2021). According to Feng (2021), the structure of a zero-inflated model is given by:

$$P(Y = y_i | \boldsymbol{\theta}) = \begin{cases} v + (1-v)p(y_i = 0 | \boldsymbol{\theta}), & y_i = 0, \\ (1-v)p(y_i | \boldsymbol{\theta}), & y_i > 0, \end{cases}$$

which is formed by a degenerated distribution at zero and a nontruncated count distribution, namely,  $p(y_i | \boldsymbol{\theta})$ , with parameter vector  $\boldsymbol{\theta}$ , where  $v \in (0, 1)$  is the probability of structural zeros.

In a hurdle, a binomial model governs the binary event when the response takes a zero or a positive value (MULLAHY, 1986). If the event is positive, the barrier is crossed, and the process is modelled by a conditional distribution that is truncated at zero (MULLAHY, 1986). According to Feng (2021), the general structure of a hurdle model is given by:

$$P(Y = y_i | \boldsymbol{\theta}) = \begin{cases} v, & y_i = 0, \\ (1-v) \frac{p(y_i | \boldsymbol{\theta})}{1 - p(y_i = 0 | \boldsymbol{\theta})}, & y_i > 0, \end{cases} \quad (1)$$

where  $v$  is the probability of  $y_i$  being zero,  $p(y_i | \boldsymbol{\theta})$  is the probability function of a count distribution and  $\boldsymbol{\theta}$  is the parameter vector of  $p(y_i | \boldsymbol{\theta})$ .

In equation (1), the first structure fits the occurrence of zeros with probability  $v$ , and the other is responsible for modelling when zeros do not occur. Note that in a hurdle, the zeros are not differentiated. In fact, events that result in zero are seen as coming from a structural source (ZUUR *et al.*, 2009).

There are several possibilities to specify the conditional distribution of the positive values and the binary event in a hurdle (MULLAHY, 1986). One of them is to consider that a binary event is modelled by a binomial distribution and the positive values follow a Poisson distribution, resulting in the zero-adjusted Poisson (ZAP) model (HASHIM; HASHIM; SHIKER, 2021).

Additional topics, such as the analysis and specification of the linear predictor in zero-inflated or hurdle models, can be seen in Zuur *et al.* (2009) and Zuo *et al.* (2021). For the analysis of count time series using zero-altered models, we define the linear predictor inserting a dependence structure below.

### Zero-altered models for count time series

Suppose  $Y$  is an equally spaced time series indexed at time  $t$ ,  $t = \{1, \dots, n\}$ . Define  $\mathcal{F}_{t-1}$  as the set of previous information until  $t-1$ , which is given by  $\mathcal{F}_{t-1} = \{y_1, y_2, \dots, y_{t-1}, \mu_1, \mu_2, \dots, \mu_{t-1}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{t-1}\}$ , where  $\mathbf{x}_t$  is the parameter vector that contains  $r$  explanatory variables,  $\mathbf{x}_t = (\mathbf{x}_{t1}, \dots, \mathbf{x}_{tr})^\top$ , under restriction that  $r < n$ . Additionally, suppose  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_r)^\top$  is the coefficient vector related to  $\mathbf{x}_t$ .

Let  $\Phi_p(B) = 1 - \phi_1 B^1 - \dots - \phi_p B^p$  and  $\Theta_q(B) = 1 - \theta_1 B^1 - \dots - \theta_q B^q$  be the autoregressive and moving average polynomials of orders  $p$  and  $q$ , respectively. Additionally, let  $\Phi_P(B^s) = 1 - \Phi_1 B^{s1} - \dots - \Phi_P B^{sP}$  and  $\Theta_Q(B^s) = 1 - \Theta_1 B^{s1} - \dots - \Theta_Q B^{sQ}$  be the seasonal autoregressive and seasonal moving average polynomials with orders  $P$  and  $Q$ , where  $B$  is the backshift operator,  $B^d y_t = y_{t-d}$ , and  $s$  is the length of the period.

In summary, we denote  $\Phi = (\phi_1, \dots, \phi_p; \Phi_1, \dots, \Phi_P)^\top$  as the set of autoregressive parameters and  $\Theta = (\theta_1, \dots, \theta_q; \Theta_1, \dots, \Theta_Q)^\top$  as the set of moving average parameters that compose the dependence structure.

### Zero-adjusted Poisson distribution

Supposing that  $y_t | \mathcal{F}_{t-1}$  follows a zero-adjusted Poisson distribution ZAP( $\mu_t, v$ ),  $p(y_t | \mathcal{F}_{t-1})$  can be expressed as:

$$p(y_t | \mathcal{F}_{t-1}) = v I_{(y_t=0)} + \left[ \frac{1-v}{1-e^{-\mu_t}} \right] e^{-\mu_t} \mu_t^{y_t} \frac{1}{y_t!} I_{(y_t>0)}, \quad (2)$$

when  $y_t \in \mathbb{N}$ ,  $\Omega = \{\mu_t, v | \mu_t > 0, 0 < v < 1\}$ .

In equation (2),  $I_{(\cdot)}$  represents the indicator function, and  $v$  is the exact probability of the series being zero at time  $t$ . Given the truncation of the distribution at zero,  $\mu_t$  is the conditional expectation in the situation where  $y_t > 0$  (RIGBY *et al.*, 2019).

The conditional mean and variance of the ZAP( $\mu_t, v$ ) distribution are given by  $E(y_t | \mathcal{F}_{t-1}) = \frac{(1-v)\mu_t}{1-e^{-\mu_t}}$  and  $V(y_t | \mathcal{F}_{t-1}) = \frac{(1-v)\mu_t}{1-e^{-\mu_t}} \left[ 1 + \mu_t - \frac{(1-v)\mu_t}{1-e^{-\mu_t}} \right]$ , respectively (RIGBY *et al.*, 2019).

$E(y_t | \mathcal{F}_{t-1})$  can be seen as a weighting of  $\mu_t$  by  $\frac{1-v}{1-e^{-\mu_t}}$ , which is the ratio of the complementary exact probability that  $y_t = 0$  and the probability that  $y_t > 0$  from a Poisson distribution with a mean of  $\mu_t$ .

Considering the logarithm link function, ensuring that  $\mu_t \subseteq \Omega$ , the linear predictor based on the GARMA(p, q) class written using the  $B$  operator is given by equation (3):

$$\begin{aligned} \log(\mu_t) &= \Phi_p(B)\Phi_q(B^s) \left[ \mathbf{x}_t^\top \boldsymbol{\beta} - \log(y_t) \right] \\ &\quad - \Theta_q(B)\Theta_\ell(B^s) [\log(y_t) - \log(\mu_t)] \\ &\quad + \log\left(\frac{y_t^2}{\mu_t}\right). \end{aligned} \quad (3)$$

It is necessary to consider a threshold ( $c$ ) that guarantees the existence of the link function at zero. A possibility is to replace  $y_{t-j}$  by  $y_{t-j}^* = \max(y_{t-j}, c)$ ,  $c \in (0, 1)$ .

Denoting  $L(\boldsymbol{\beta}, \Phi, \Theta, \nu | Y)$  as the approximated likelihood function of the ZAP( $\mu_t, \nu$ ) model conditioned to  $m = \max(p, q, sP, sQ)$  observations, we have:

$$L(\boldsymbol{\beta}, \Phi, \Theta, \nu | Y) \approx \prod_{t=m+1}^n p(y_t | \mathcal{F}_{t-1}). \quad (4)$$

Given the conditioning of the likelihood function to the first  $m$  terms, we consider that the first terms of the error are zero, as described in the studies of Benjamin, Rigby and Stasinopoulos (2003), Rocha and Cribari-Neto (2009).

In the situation where  $y_t > 0$ , the process is assumed to be equidispersed and modelled by a Poisson distribution, i.e.,  $E(y_t | \mathcal{F}_{t-1}) = V(y_t | \mathcal{F}_{t-1}) = \mu_t$ . The following models are over dispersion alternatives.

#### Zero-adjusted Negative binomial distribution

We assume that  $y_t | \mathcal{F}_{t-1}$  follows a zero-adjusted Negative binomial distribution ZANBI( $\mu_t, \sigma, \nu$ ), with a conditional density given by:

$$\begin{aligned} p(y_t | \mathcal{F}_{t-1}) &= \nu \mathbf{I}_{(y_t=0)} + \frac{(1-\nu)\mu_t}{[1-(1+\mu_t\sigma)^{-\frac{1}{\sigma}}]} \\ &\quad \times \frac{\Gamma(y_t + \frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(y_t + 1)} \left( \frac{\sigma\mu_t}{1+\sigma\mu_t} \right)^{y_t} \\ &\quad \times \left( \frac{1}{1+\sigma\mu_t} \right)^{\frac{1}{\sigma}} \mathbf{I}_{(y_t>0)}, \end{aligned} \quad (5)$$

$\Omega = \{\mu_t, \sigma, \nu | \mu_t, \sigma > 0, 0 < \nu < 1\}$  for  $y_t \in \mathbb{N}$ .

The conditional mean and variance are given by  $E(y_t | \mathcal{F}_{t-1}) = \frac{(1-\nu)\mu_t}{[1-(1+\mu_t\sigma)^{-\frac{1}{\sigma}}]}$  and  $V(y_t | \mathcal{F}_{t-1}) =$

$$\frac{(1-\nu)\mu_t}{[1-(1+\mu_t\sigma)^{-\frac{1}{\sigma}}]} \left[ 1 + \mu_t \left( 1 + \sigma - \frac{(1-\nu)}{[1-(1+\mu_t\sigma)^{-\frac{1}{\sigma}}]} \right) \right],$$

respectively, where  $\nu$  is the exact probability of the series being equal to zero. The  $\mu_t$  and  $\sigma$  parameters are the mean and dispersion of the process in the situation where  $y_t > 0$ , respectively (RIGBY *et al.*, 2019).

We adopt the same structure in the linear predictor described by equation (3). However, the approximate likelihood function is written according to equation (6).

$$L(\boldsymbol{\beta}, \Phi, \Theta, \nu, \sigma | Y) \approx \prod_{t=m+1}^n p(y_t | \mathcal{F}_{t-1}). \quad (6)$$

The assumptions that the first  $m$  error terms are zero follow as detailed in the ZAP( $\mu_t, \nu$ ) model. Generally, the process is governed by a truncation at zero and a Negative binomial distribution with a mean of  $\mu_t$  and dispersion of  $\sigma$  in the condition where  $y_t > 0$ .

#### Zero-adjusted Poisson inverse Gaussian distribution

Assuming that  $y_t | \mathcal{F}_{t-1}$  follows a zero-adjusted Poisson inverse Gaussian distribution, ZAPIG( $\mu_t, \sigma, \nu$ ), the conditional density function, is given by equation (7):

$$\begin{aligned} p(y_t | \mathcal{F}_{t-1}) &= \nu \mathbf{I}_{(y_t=0)} \\ &\quad + \frac{1-\nu}{1-e^{\frac{1}{\sigma}-\alpha}} \left( \frac{2\alpha}{\pi} \right)^{\frac{1}{2}} \mu_t^{y_t} e^{\frac{1}{\sigma}} \left( K_{y_t-\frac{1}{2}}(\alpha) \right) \\ &\quad \times \frac{1}{(\alpha\sigma)^{y_t} y_t!} \mathbf{I}_{(y_t>0)}, \end{aligned} \quad (7)$$

$\Omega = \{\mu_t, \sigma, \nu | \mu_t, \sigma > 0, 0 < \nu < 1\}$  for  $y_t \in \mathbb{N}$ .

In equation (7),  $\alpha^2 = \frac{1}{\sigma^2} + \frac{2\mu_t}{\sigma}$  and  $K_\lambda(T) = \frac{1}{2} \int_0^\infty x^{\lambda-1} \exp\{-\frac{T}{2}(x+x^{-1})\} dx$ , where  $K(\cdot)$  is the modified Bessel function of the third kind<sup>1</sup>. The conditional mean and variance are  $E(y_t | \mathcal{F}_{t-1}) = \frac{(1-\nu)\mu_t}{1-e^{\frac{1}{\sigma}-\alpha}}$  and  $V(y_t | \mathcal{F}_{t-1}) = \left[ \frac{(1-\nu)\mu_t}{1-e^{\frac{1}{\sigma}-\alpha}} \right] \left[ 1 + \mu_t \left( 1 + \sigma - \frac{1-\nu}{1-e^{\frac{1}{\sigma}-\alpha}} \right) \right]$ , respectively.

It is important to note that  $\nu$  is the exact probability of the series being zero, and it is  $p(y_t = 0 | \mathcal{F}_{t-1}) = \nu, \forall t$ , as well as in the ZAP( $\mu_t, \nu$ ) and ZANBI( $\mu_t, \sigma, \nu$ ) models. The  $\mu_t$  and  $\sigma$  parameters are related to the mean and dispersion of the series in the situation where  $y_t > 0$  (RIGBY *et al.*, 2019).

We also considered the linear predictor shown in equation (3), ensuring that  $\mu_t \subseteq \Omega$ . The approximated likelihood function is analogous to equation (6) and the assumptions that the first  $m$  error terms are zero follow as detailed in the ZAP( $\mu_t, \nu$ ) model.

<sup>1</sup> Modified Bessel functions are particular forms of Bessel functions, which are solutions of the Bessel equation, being related to the Poisson process and applied in the analysis of spherical distributions (KORENEV, 2002; ROBERT, 1990). For details, see Korenev (2002).

### Bayesian analysis

We consider  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \nu)^\top$  in the zero-adjusted Poisson model and  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \nu, \sigma)^\top$  in the zero-adjusted Negative binomial and zero-adjusted Poisson inverse Gaussian models. In this way, the resulting posterior, namely,  $\pi(\boldsymbol{\theta} | Y)$ , is proportional to  $L(\boldsymbol{\theta} | Y) \times \pi_0(\boldsymbol{\theta})$ , where  $L(\boldsymbol{\theta} | Y)$  is the approximated likelihood function and  $\pi_0(\boldsymbol{\theta})$  is the joint prior of  $\boldsymbol{\theta}$ .

We consider noninformative priors for  $\boldsymbol{\theta}$  in all models. For the parameters associated with the explanatory variables, each component of  $\boldsymbol{\beta}$  is normally distributed. It is given as:

$$p(\beta_j) \propto \exp \left[ -\frac{1}{2} \left( \frac{\beta_j - \mu_j}{\tau_j} \right)^2 \right],$$

where  $\beta_j \in (-\infty, \infty)$ ,  $j = \{1, \dots, r\}$ , and the hyperparameters are  $\mu_j = 0$  and  $\tau_j = 100$ . The same structure of priors used for  $\boldsymbol{\beta}$  is adopted for  $\boldsymbol{\Phi}$  and  $\boldsymbol{\Theta}$ .

For the parameter  $\nu$ , we consider a uniform distribution on the interval  $(0, 1)$ , and for the models that contain a dispersion parameter, we use a Gamma( $a, s$ ) distribution with hyperparameters  $a = 1$  and  $s = 100^{-1}$ , with the following density:

$$p(\sigma) = \frac{\sigma^{a-1}}{s^a \Gamma(a)} \exp \left( -\frac{\sigma}{s} \right), \quad \sigma > 0.$$

Given the algebraic complexity to obtain closed forms of  $\pi(\boldsymbol{\theta} | Y)$ , the inference can be performed using Markov chain Monte Carlo (MCMC) methods, such as the Metropolis-Hastings (MH) algorithm for drawing from the joint posterior of each model (ANDRADE; ANDRADE; EHLERS, 2015).

#### Predictive density

To make forecasts of  $Y$  over a horizon  $h$ , it is necessary to define the predictive distribution, which is the distribution of  $y_{t+h}$  conditioning to all parameters and previous observations that compose the set  $\mathcal{F}_{t-1}$  (BROEMELING, 2019; SÁFADI; MORETTIN, 2003).

Combining the joint posterior  $\pi(\boldsymbol{\theta} | Y)$  with the density of the new observation,  $p(y_{t+h} | \boldsymbol{\theta}, \mathcal{F}_{t+h-1})$ , the predictive density is given by:

$$p(y_{t+h} | \mathcal{F}_{t+h-1}) = \int_{\boldsymbol{\theta} \in \Omega} p(y_{t+h} | \boldsymbol{\theta}, \mathcal{F}_{t+h-1}) \pi(\boldsymbol{\theta} | Y) d\boldsymbol{\theta}.$$

There is no closed form in the considered models. Nonetheless, we can perform a Monte Carlo approximation of the predictive, drawing  $\mathcal{N}$  samples from  $\boldsymbol{\theta}^i$ ,

$i = \{1, \dots, \mathcal{N}\}$ , as follows:

$$p(y_{t+h} | \mathcal{F}_{t+h-1}) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} p(y_{t+h} | \boldsymbol{\theta}^i, \mathcal{F}_{t+h-1}).$$

Then, the expected value of  $y_{t+h}$  is given by the following expression:

$$E(y_{t+h}) = \int_{y_{t+h} \in \Omega} y_{t+h} p(y_{t+h} | \mathcal{F}_{t+h-1}) dy_{t+h}.$$

The approximation of the expected value is performed using the conditional mean,  $E(y_t | \mathcal{F}_{t-1})$ , of each distribution. For the zero-adjusted Poisson model:

$$\hat{y}_{t+h} \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \frac{(1 - \boldsymbol{\theta}_4^i) \mu_{t+h}(\boldsymbol{\theta}_1^i, \boldsymbol{\theta}_2^i, \boldsymbol{\theta}_3^i, \mathcal{F}_{t+h-1})}{1 - e^{-\mu_{t+h}(\boldsymbol{\theta}_1^i, \boldsymbol{\theta}_2^i, \boldsymbol{\theta}_3^i, \mathcal{F}_{t+h-1})}}.$$

Similarly, for the zero-adjusted Negative binomial model:

$$\hat{y}_{t+h} \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \frac{(1 - \boldsymbol{\theta}_4^i) \mu_{t+h}(\boldsymbol{\theta}_1^i, \boldsymbol{\theta}_2^i, \boldsymbol{\theta}_3^i, \mathcal{F}_{t+h-1})}{\left\{ 1 - [1 + \mu_{t+h}(\boldsymbol{\theta}_1^i, \boldsymbol{\theta}_2^i, \boldsymbol{\theta}_3^i, \mathcal{F}_{t+h-1}) \boldsymbol{\theta}_5^i]^{-\frac{1}{\boldsymbol{\theta}_5^i}} \right\}}.$$

Considering the zero-adjusted Poisson inverse Gaussian model, we have:

$$\hat{y}_{t+h} \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \frac{(1 - \boldsymbol{\theta}_4^i) \mu_{t+h}(\boldsymbol{\theta}_1^i, \boldsymbol{\theta}_2^i, \boldsymbol{\theta}_3^i, \mathcal{F}_{t+h-1})}{\left[ 1 - e^{\left( \frac{1}{\boldsymbol{\theta}_5^i} - \alpha \right)} \right]}.$$

### A simulation study

We carry out a simulation study considering the ZAP-AR(1), ZANBI-AR(1) and ZAPIG-AR(1) models, which are expressed as follows:

$$\begin{aligned} y_t | \mathcal{F}_{t-1} &\sim \text{ZAP} \{ \mu_t, \nu \} \\ y_t | \mathcal{F}_{t-1} &\sim \text{ZANBI} \{ \mu_t, \nu, \sigma \} \\ y_t | \mathcal{F}_{t-1} &\sim \text{ZAPIG} \{ \mu_t, \nu, \sigma \}, \end{aligned}$$

where  $\log(\mu_t) = (1 - \phi_1 B^1) [\mathbf{x}_t^\top \boldsymbol{\beta} - \log(y_t)] + \log(y_t)$ . We evaluate the scenarios presented in Table 1, with  $\boldsymbol{\beta} = (\beta_1)^\top$  and  $\mathbf{x}_t = (\mathbf{x}_{t1})^\top$ , where  $\mathbf{x}_{t1} = (1, 1, \dots, 1)^\top$ . We establish time series with 10% and 40% of counts equal to zero and correlations of 0.30 and -0.80. For models with dispersion parameter, we fix  $\sigma$  at 0.10.

**Table 1** – Parameter settings for the ZAP-AR(1), ZANBI-AR(1) and ZAPIG-AR(1) models.

Scenario	n	$\beta_1$	$\phi_1$	$\nu$	$\sigma$
I	100	2.00	0.30	{0.10; 0.40}	0.10
II	100	2.00	-0.80	{0.10; 0.40}	0.10

Source: The authors.

We used the MH algorithm for drawing from the joint posterior, which is described in the appendix, with the *MHadaptive* package from Chivers (2015), and each model was replicated  $w = 1,000$  times. The pseudorandom series were generated with the *gamlss.dist* package from Stasinopoulos and Rigby (2020) using the inverse transformed method. The burn-in, thinness and number of samples needed to obtain convergence to the stationary distribution were defined in a pilot study.

We formally verified the convergence in each replicated model using the *HW* (HEIDELBERGER; WELCH, 1983), *G* (GEWEKE, 1991) and the dependence factor (*I*) (RAFTERY; LEWIS, 1992) diagnoses. For the *HW*, we fixed  $\alpha = 5\%$ ; the absolute value of *G* was compared with the quantile  $Z_{1-\frac{\alpha}{2}}$ , and we checked if  $I \rightarrow 1$ . By meeting these specifications, the process was considered convergent.

#### Simulation results

Table 2 displays the simulation results of the first scenario, where  $\phi_1 = 0.30$ . We observed that there was an increase in the standard deviation of the parameter  $\beta_1$  with an increase in the proportion of zeros ( $\nu$ ). The greatest increase in the standard deviation was related to the intercept of the ZAP-AR(1) model, which increased by 85.93% in the case where  $\nu = 0.40$ .

According to the scenario displayed in Table 2, we verified that the mean and mode estimates of  $\nu$  remain identical up to the third decimal place. However, there was a reduction in the CB and CE metrics of  $\nu$  when the proportion of zeros increased.

For models with dispersion parameter, better results were obtained using the estimates based on the mode compared with those using the posterior mean. We also observed an increase in the standard deviation of  $\sigma$  when increasing  $\nu$ , mainly in the ZAPIG-AR(1) model, where some details about the settings of the MH algorithm are available in the Appendix, presented in the *Additional Simulation Results*. Also, in the appendix, the simulation results of the second scenario are displayed and the MH settings are available. Similar results were obtained in the second scenario when compared to the first scenario, mainly when evaluating the dispersion parameter estimates based on the posterior mode in the ZANBI-AR(1) and ZAPIG-AR(1) models. The results of the evaluated scenarios indicated good inference properties, which can be verified in the CB and CE metrics. All the CE values are approximately one, and great acceptance rates were obtained using the MH algorithm.

152

#### Application: mortality from dengue haemorrhagic fever

We analysed the monthly mortality series due to dengue haemorrhagic fever (ICD-A91) of children under nine years old in Brazil. Our data were provided by Brasil ([2022]) and covered the period from January 2013 to July 2020. The behaviour of the data over time is shown in Figure 1(a), and the respective bar graph is presented in Figure 1(b).

In the studied period, 136 deaths were reported, resulting in an average of 1.53 deaths per month, with a median equal to one. The proportion of months without occurrence of deaths was 33.71%, estimated by  $\hat{p} = \frac{30}{89}$ , where 89 is the number of months. Considering the test for excess zeros proposed by Broek (1995), based on the Poisson distribution, the data were zero inflated ( $p$ -value  $< 0.001$ ). The same result was found using the *check\_zeroinflation()* function from Lüdecke *et al.* (2021).

To model the series, we considered the models presented in equation (8), with  $\boldsymbol{\beta} = \beta_1$  and  $\mathbf{x}_t^\top = (\mathbf{x}_{t1})^\top$ ,  $\mathbf{x}_{t1} = (1, 1, \dots, 1)^\top$ . We used the MH algorithm with a burn-in period of 1,000 and thin = 15, resulting in a final sample of 10,000.

$$\begin{aligned} y_t | \mathcal{F}_{t-1} &\sim \text{ZAP}\{\mu_t, \nu\} \\ y_t | \mathcal{F}_{t-1} &\sim \text{ZANBI}\{\mu_t, \nu, \sigma\} \\ y_t | \mathcal{F}_{t-1} &\sim \text{ZAPIG}\{\mu_t, \nu, \sigma\}, \end{aligned} \quad (8)$$

where  $\log(\mu_t) = (1 - \phi_1 B^1)(1 - \Phi_1 B^{12})[\beta_1 - \log(y_t)] + \log(y_t)$ .

Table 3 lists the estimation results for the models expressed in equation (8), including the credibility intervals and the deviance information criterion (DIC). The  $\Phi_1$  and  $\beta_1$  parameters were removed from the ZANBI( $\mu_t, \sigma, \nu$ ) model because they were not significant. The same situation occurred with the  $\beta_1$  parameter of the ZAPIG( $\mu_t, \sigma, \nu$ ). We also presented the dependence factor to evaluate the convergence of the chains.

The *Box-Pierce* (BOX; PIERCE, 1970) test indicated that the residuals of the models presented in Table 3 are noncorrelated, with  $p$ -values of 0.89, 0.11 and 0.90 for ZAP-SAR(1)(1)<sub>12</sub>, ZANBI-AR(1) and ZAPIG-SAR(1)(1)<sub>12</sub>, respectively, evaluated up to lag 20.

The lowest DIC was obtained using the ZAP-SAR(1)(1)<sub>12</sub> model, and the process can be written according to the following scheme:

$$\begin{aligned} y_t | \mathcal{F}_{t-1} &\sim \text{ZAP}\{\log(\mu_t) = (1 - 0.381B^1)(1 - 0.253B^{12}) \\ &\quad \times [1.062 - \log(y_t)] + \log(y_t), \nu = 0.400\}. \end{aligned}$$



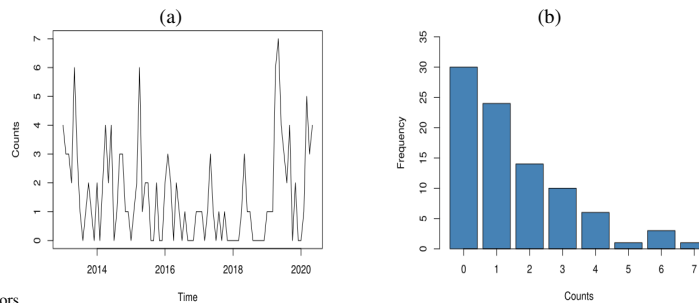
Count time series with excess zeros: A Bayesian perspective using zero-adjusted distributions

**Table 2** – Simulation results according to the first scenario, considering  $w = 1,000$  replicates of each model.

Model	n	Parameter	Real	Mean	Mode	SD	CB	CE
ZAP-AR(1)	100	$\beta_1$	2.000	1.998	1.995	0.064	0.025	1.000
		$\phi_1$	0.300	0.308	0.305	0.053	0.140	1.012
		$\nu$	0.100	0.108	0.101	0.029	0.236	1.035
	100	$\beta_1$	2.000	2.004	1.996	0.119	0.046	1.000
		$\phi_1$	0.300	0.306	0.304	0.045	0.117	1.007
		$\nu$	0.400	0.404	0.402	0.049	0.096	1.003
ZANBI-AR(1)	100	$\beta_1$	2.000	2.004	1.997	0.090	0.036	1.000
		$\phi_1$	0.300	0.313	0.309	0.057	0.154	1.024
		$\nu$	0.100	0.108	0.101	0.029	0.236	1.038
	100	$\sigma$	0.100	0.124	0.108	0.047	0.404	1.117
		$\beta_1$	2.000	2.022	2.002	0.151	0.060	1.010
		$\phi_1$	0.300	0.313	0.309	0.050	0.136	1.031
ZAPIG-AR(1)	100	$\nu$	0.400	0.395	0.393	0.047	0.095	1.005
		$\sigma$	0.100	0.165	0.119	0.072	0.730	1.356
		$\beta_1$	2.000	2.006	1.997	0.088	0.035	1.002
	100	$\phi_1$	0.300	0.312	0.308	0.063	0.169	1.017
		$\nu$	0.100	0.107	0.100	0.029	0.233	1.027
		$\sigma$	0.100	0.127	0.108	0.050	0.433	1.135
100	$\beta_1$	2.000	2.035	2.003	0.158	0.065	1.024	
	$\phi_1$	0.300	0.311	0.306	0.054	0.145	1.021	
	$\nu$	0.400	0.398	0.396	0.047	0.093	1.021	
100	$\sigma$	0.100	0.182	0.126	0.095	0.879	1.319	

Source: The authors.

**Figure 1** – Data over time: (a) monthly number of deaths of individuals under nine years of age due to A91 in Brazil; (b) bar graph of the data presented in (a).



Source: The authors.

The estimate of  $\nu$  indicates that the probability of nonoccurrence of any death of individuals aged 0 to 9 years due to A91 in a given month  $t$  is 40.00% and varies between 28.60% and 52.10% according to the credibility interval. When considering  $1 - \nu$ , the exact probability of at least one death is 60.00%.

The average number of deaths in a given month  $t$ , conditioned to  $y_t > 0$ , can be estimated using the parameter  $\mu_t$ , i.e.,  $\log(\mu_t) = (1 - 0.381B^1)(1 - 0.253B^{12})[1.062 - \log(y_t)] + \log(y_t)$ .

In Figure 2, we show the forecasts for June and July 2020, where the grey areas represent the regions of high

density at the 95% credibility level. The estimates point to 2.05 deaths in June and 1.63 in July. Compared with the real values of three and one death in June and July, respectively, the best forecast was for the month of July. Figure 2(c) also displays the fitted values and the forecasts for June and July 2020, with the 95% credibility intervals.

Based on the predictive density, the probability  $P(\hat{y}_{Jun} = 3 | \mathcal{F}_{t-1})$  is 13.47%, with  $CI_{95\%} = (10.50\%; 16.60\%)$ . For July,  $P(\hat{y}_{Jul} = 1 | \mathcal{F}_{t-1})$  is 13.76%, with  $CI_{95\%} = (7.00\%; 20.60\%)$ . In terms of accuracy, a mean absolute percent error (MAPE) equal to 47.33% was obtained using the ZAP-SAR(1)(1)<sub>12</sub> model.

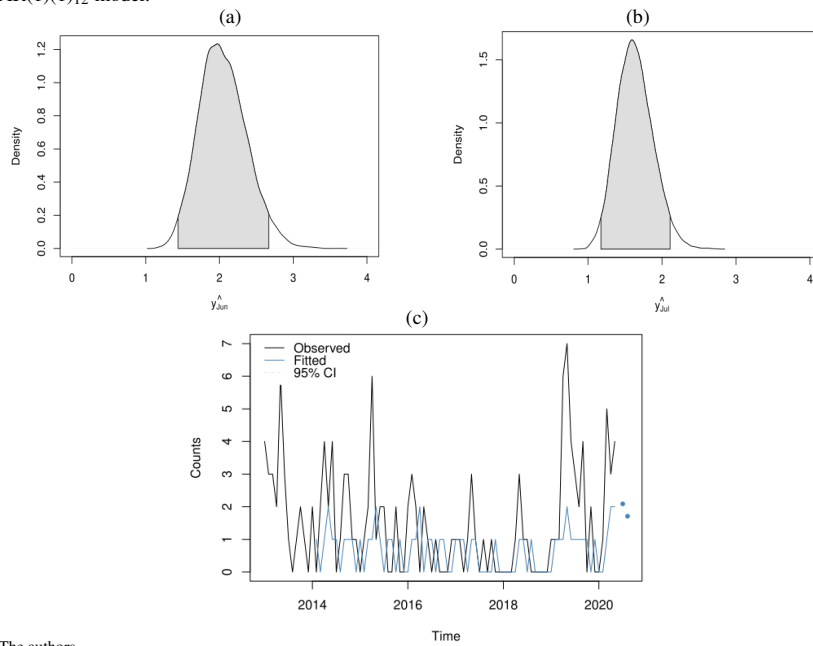
Pala, L. O. O.; Carvalho, M. M.; Sáfiadi, T.

**Table 3** – Parameter estimates for the models defined in equation (8) and fit to the dengue data (A91) for children under nine years old.

Model	Parameter	Mean	SD	HPD (95%)		I	DIC
				$L_l$	$L_u$		
ZAP-SAR(1)(1) <sub>12</sub>	$\beta_1$	1.062	0.330	0.420	1.726	1.350	569.030
	$\phi_1$	0.381	0.139	0.124	0.663	1.100	
	$\Phi_1$	0.253	0.120	0.019	0.483	1.090	
	$\nu$	0.400	0.060	0.286	0.521	1.050	
ZANBI-AR(1)	$\phi_1$	0.478	0.143	0.193	0.753	1.080	725.954
	$\nu$	0.345	0.050	0.245	0.439	1.030	
	$\sigma$	0.957	0.433	0.248	1.825	1.010	
ZAPIG-SAR(1)(1) <sub>12</sub>	$\phi_1$	0.446	0.173	0.121	0.800	1.170	613.481
	$\Phi_1$	0.372	0.175	0.035	0.718	1.100	
	$\nu$	0.372	0.054	0.266	0.476	1.050	
	$\sigma$	0.784	0.524	0.019	1.775	0.976	

Source: The authors.

**Figure 2** – Density of the expected number of deaths due to A91: (a) June; (b) July; (c) fitted values according to the ZAP-SAR(1)(1)<sub>12</sub> model.



Source: The authors.

In Table 4, we show the probability of extreme values, such as 8, 10 and 12, which is a form of analysis that allows evaluating periods of outbreak and increase in the number of deaths (YANG; ZAMBA; CAVANAUGH, 2013). According to these results, the probability that the number of deaths is greater than or equal to 12 is 0.018% in June and 0.001% in July 2020.

**Table 4** – Probabilities of occurrence of atypical counts in the months of June and July 2020 based on the predictive density.

Month	$P(\hat{y}_t \geq 8   \cdot)$	$P(\hat{y}_t \geq 10   \cdot)$	$P(\hat{y}_t \geq 12   \cdot)$
June	0.014	0.001	<0.001
July	0.003	<0.001	<0.001

Source: The authors.

**Conclusions**

In this work, the Poisson, Negative binomial and, Poisson inverse Gaussian zero-adjusted distributions were studied for modelling count time series. A Bayesian analysis was adopted, and its performance was evaluated using a simulation study.

When analysing mortality due to dengue haemorrhagic fever in Brazil, we verified the usefulness of the zero-adjusted distributions, which allow considering characteristics such as the excess of zeros and estimating the exact probabilities that the process assumes the zero value at a given instant of time.

These contributions allow the analysis of time series, such as the number of disease cases in populations, monitoring and predicting periods and probabilities of outbreaks over time, as discussed in Yang, Zamba and Cavanaugh (2013) and Sathish, Mukhopadhyay and Tiwari (2021).

Suggestions for future work are the study of higher ARMA(p, q) orders, the increase in the proportion of zeros and the implementation and comparison of other algorithms for drawing from the joint posterior, such as Hamiltonian Monte Carlo (HMC) algorithm (DUANE *et al.*, 1987).

**Acknowledgments**

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001*.

**Appendices**

*Joint posteriors of the simulated models*

This appendix presents the joint posterior of the simulated models. Considering the ZAP-AR(1) model, we have:

$$\pi(\boldsymbol{\theta} | Y) \propto \prod_{t=m+1}^n \left\{ vI_{(y_t=0)} + \left[ \frac{1-v}{1-e^{-v}} \right] \frac{e^{-\mu_t} \mu_t^{y_t}}{y_t!} I_{(y_t>0)} \right\} \times \exp \left\{ -\frac{(\beta_1^2 + \phi_1^2)}{200} \right\},$$

where  $\mu_t = \exp\{[1 - \phi_1 B^1][\beta_1 - \log(y_t)] + \log(y_t)\}$ .

For the ZANBI-AR(1) model, we have:

$$\pi(\boldsymbol{\theta} | Y) \propto \prod_{t=m+1}^n \left\{ vI_{(y_t=0)} + \frac{(1-v)\mu_t}{[1 - (1 + \mu_t \sigma)^{\frac{1}{\sigma}}]} \times \frac{\Gamma(y_t + \frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(y_t + 1)} \left( \frac{\sigma \mu_t}{1 + \sigma \mu_t} \right)^{y_t} \left( \frac{1}{1 + \sigma \mu_t} \right)^{\frac{1}{\sigma}} \times I_{(y_t>0)} \right\} \times \exp \left[ -\frac{(\beta_1^2 + \phi_1^2)}{200} - \frac{\sigma}{s} \right] \frac{\sigma^{a-1}}{s^a \Gamma(a)},$$

where  $\mu_t = \exp\{[1 - \phi_1 B^1][\beta_1 - \log(y_t)] + \log(y_t)\}$ .

For the ZAPIG-AR(1) model, we have:

$$\pi(\boldsymbol{\theta} | Y) \propto \prod_{t=m+1}^n \left\{ vI_{(y_t=0)} + \left[ \frac{1-v}{1-e^{\frac{1}{\sigma}-\alpha}} \right] \times \left( \frac{2\alpha}{\pi} \right)^{\frac{1}{2}} \mu_t^{y_t} e^{\frac{1}{\sigma}} \left( K_{y_t-\frac{1}{2}}(\alpha) \right) \frac{1}{(\alpha \sigma)^{y_t} y_t!} I_{(y_t>0)} \right\} \times \left\{ \exp \left[ -\frac{(\beta_1^2 + \phi_1^2)}{200} - \frac{\sigma}{s} \right] \frac{\sigma^{a-1}}{s^a \Gamma(a)} \right\},$$

where  $\mu_t = \exp\{[1 - \phi_1 B^1][\beta_1 - \log(y_t)] + \log(y_t)\}$ .

*Joint posterior distributions applied to the dengue data*

Considering the time series, the joint posterior of the ZAP-SAR(1)(1)<sub>12</sub> model is given by:

$$\pi(\boldsymbol{\theta} | Y) \propto \prod_{t=m+1}^n \left\{ vI_{(y_t=0)} + \left[ \frac{1-v}{1-e^{-v}} \right] \frac{e^{-\mu_t} \mu_t^{y_t}}{y_t!} I_{(y_t>0)} \right\} \times \exp \left\{ -\frac{(\beta_1^2 + \phi_1^2 + \Phi_1^2)}{200} \right\},$$

where  $\mu_t = \exp\{[1 - \phi_1 B^1][1 - \Phi_1 B^{12}][\beta_1 - \log(y_t)] + \log(y_t)\}$ . Considering the ZANBI-AR(1) model, we have:

$$\pi(\boldsymbol{\theta} | Y) \propto \prod_{t=m+1}^n \left\{ vI_{(y_t=0)} + \frac{(1-v)\mu_t}{[1 - (1 + \mu_t \sigma)^{\frac{1}{\sigma}}]} \times \frac{\Gamma(y_t + \frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(y_t + 1)} \left( \frac{\sigma \mu_t}{1 + \sigma \mu_t} \right)^{y_t} \left( \frac{1}{1 + \sigma \mu_t} \right)^{\frac{1}{\sigma}} \times I_{(y_t>0)} \right\} \times \left\{ \exp \left[ -\frac{\phi_1^2}{200} - \frac{\sigma}{s} \right] \frac{\sigma^{a-1}}{s^a \Gamma(a)} \right\},$$

where  $\mu_t = \exp\{[1 - \phi_1 B^1][-\log(y_t)] + \log(y_t)\}$ . For the ZAPIG-SAR(1)(1)<sub>12</sub> model, we have:

$$\pi(\boldsymbol{\theta} | Y) \propto \prod_{t=m+1}^n \left\{ vI_{(y_t=0)} + \left[ \frac{1-v}{1-e^{\frac{1}{\sigma}-\alpha}} \right] \times \left( \frac{2\alpha}{\pi} \right)^{\frac{1}{2}} \mu_t^{y_t} e^{\frac{1}{\sigma}} \left( K_{y_t-\frac{1}{2}}(\alpha) \right) \frac{1}{(\alpha \sigma)^{y_t} y_t!} I_{(y_t>0)} \right\} \times \left\{ \exp \left[ -\frac{(\phi_1^2 + \Phi_1^2)}{200} - \frac{\sigma}{s} \right] \frac{\sigma^{a-1}}{s^a \Gamma(a)} \right\},$$

being  $\mu_t = \exp\{[1 - \phi_1 B^1][1 - \Phi_1 B^{12}][-\log(y_t)] + \log(y_t)\}$ .

*Additional simulation results*

Table 5 displays the simulation results for the second scenario. In Tables 6 and 7, we present the settings of the MH algorithm and the convergence analysis.

Pala, L. O. O.; Carvalho, M. M.; Sáfadi, T.

**Table 5** – Second scenario simulation results based on  $w = 1,000$  replicated models.

Model	n	Parameter	Real	Mean	Mode	SD	CB	CE
ZAP-AR(1)	100	$\beta_1$	2.000	2.000	2.000	0.024	0.010	1.000
		$\phi_1$	-0.800	-0.800	-0.800	0.012	0.012	1.000
		$\nu$	0.100	0.108	0.101	0.029	0.234	1.035
	100	$\beta_1$	2.000	1.999	1.998	0.042	0.017	1.000
		$\phi_1$	-0.800	-0.801	-0.801	0.018	0.018	1.001
		$\nu$	0.400	0.400	0.398	0.048	0.096	1.000
ZANBI-AR(1)	100	$\beta_1$	2.000	1.997	1.997	0.035	0.014	1.003
		$\phi_1$	-0.800	-0.802	-0.801	0.029	0.029	1.002
		$\nu$	0.100	0.110	0.103	0.030	0.251	1.081
	100	$\sigma$	0.100	0.113	0.102	0.032	0.267	1.052
		$\beta_1$	2.000	1.992	1.991	0.069	0.027	1.007
		$\phi_1$	-0.800	-0.805	-0.803	0.034	0.033	1.008
ZAPIG-AR(1)	100	$\nu$	0.400	0.403	0.402	0.046	0.092	1.002
		$\sigma$	0.100	0.113	0.101	0.030	0.251	1.098
		$\beta_1$	2.000	1.997	1.996	0.036	0.014	1.003
	100	$\phi_1$	-0.800	-0.804	-0.803	0.029	0.029	1.008
		$\nu$	0.100	0.108	0.101	0.029	0.234	1.037
		$\sigma$	0.100	0.113	0.100	0.032	0.269	1.077
ZAPIG-AR(1)	100	$\beta_1$	2.000	1.997	1.996	0.067	0.026	1.001
		$\phi_1$	-0.800	-0.804	-0.802	0.034	0.033	1.006
		$\nu$	0.400	0.402	0.400	0.048	0.095	1.000
	100	$\sigma$	0.100	0.116	0.102	0.032	0.274	1.115

Source: The authors.

**Table 6** – Settings used in MH and convergence analysis for the first scenario, where  $\overline{AC}$  is the probability acceptance,  $\bar{I}$  is the mean of the dependence factor,  $\bar{G}$  is the mean of the absolute value of G, and  $\overline{HW}$  is the mean of the p-value of HW test.

Model	n	Samples	Burn-in	Thin	$\overline{AC}$	$\theta$	$\bar{I}$	$\bar{G}$	$\overline{HW}$
ZAP-AR(1)	100	5,000	1,000	7	0.48	$\beta_1$	1.24	0.00	0.50
						$\phi_1$	1.21	0.05	0.51
						$\nu$	1.13	0.03	0.51
	100	5,000	1,000	8	0.45	$\beta_1$	1.22	0.02	0.50
						$\phi_1$	1.23	0.02	0.50
						$\nu$	1.22	0.05	0.49
ZANBI-AR(1)	100	5,000	1,000	11	0.36	$\beta_1$	1.20	0.04	0.49
						$\phi_1$	1.19	0.04	0.47
						$\nu$	1.10	0.07	0.49
	100	6,000	1,000	11	0.34	$\sigma$	1.08	0.04	0.50
						$\beta_1$	1.25	0.06	0.51
						$\phi_1$	1.25	0.06	0.50
ZAPIG-AR(1)	100	5,000	1,000	11	0.36	$\nu$	1.24	0.06	0.49
						$\sigma$	1.12	0.07	0.50
						$\beta_1$	1.21	0.08	0.50
	100	5,000	1,000	11	0.36	$\phi_1$	1.20	0.10	0.49
						$\nu$	1.12	0.13	0.48
						$\sigma$	1.09	0.10	0.48
100	5,000	1,000	11	0.34	$\beta_1$	1.21	0.05	0.50	
					$\phi_1$	1.26	0.10	0.50	
					$\nu$	1.24	0.00	0.48	
100	5,000	1,000	11	0.34	$\sigma$	1.13	0.04	0.51	

Source: The authors.

**Table 7** – Settings used in MH and convergence analysis for the second scenario, where  $\overline{AC}$  is the probability acceptance,  $\bar{I}$  is the mean of the dependence factor,  $\bar{G}$  is the mean of the absolute value of  $G$ , and  $\overline{HW}$  is the mean of the p-value of HW test.

Model	n	Samples	Burn-in	Thin	$\overline{AC}$	$\theta$	$\bar{I}$	$\bar{G}$	$\overline{HW}$
ZAP-AR(1)	100	5,000	1,000	8	0.45	$\beta_1$	1.25	0.08	0.50
						$\phi_1$	1.25	0.07	0.50
						$\nu$	1.13	0.14	0.49
	100	5,000	1,000	8	0.45	$\beta_1$	1.24	0.13	0.49
						$\phi_1$	1.26	0.13	0.49
						$\nu$	1.22	0.14	0.49
ZANBI-AR(1)	100	5,000	1,000	9	0.36	$\beta_1$	1.26	0.06	0.51
						$\phi_1$	1.30	0.08	0.50
						$\nu$	1.15	0.08	0.49
	100	5,000	1,000	10	0.36	$\sigma$	1.12	0.06	0.50
						$\beta_1$	1.22	0.00	0.48
						$\phi_1$	1.25	0.00	0.49
ZAPIG-AR(1)	100	6,000	1,000	12	0.35	$\nu$	1.20	0.03	0.49
						$\sigma$	1.08	0.13	0.51
						$\beta_1$	1.16	0.06	0.50
	100	6,000	1,000	12	0.35	$\phi_1$	1.20	0.01	0.50
						$\nu$	1.07	0.09	0.50
						$\sigma$	1.05	0.14	0.51
100	6,000	1,000	12	0.35	$\beta_1$	1.17	0.07	0.49	
					$\phi_1$	1.20	0.05	0.50	
					$\nu$	1.15	0.06	0.48	
						$\sigma$	1.05	0.16	0.50

Source: The authors.

## References

- ANDRADE, B.; ANDRADE, M.; EHLERS, R. Bayesian GARMA models for count data. *Communications in statistics: case studies, data analysis and applications*, Philadelphia, v. 1, n. 4, p. 192-205, 2015. DOI: <https://doi.org/10.1080/23737484.2016.1190307>.
- ALQAWBA, M.; DIAWARA, N.; CHAGANTY, N. Zero-inflated count time series models using Gaussian copula. *Sequential Analysis*, New York, v. 38, n. 3, p. 342-357, 2019. DOI: <https://doi.org/10.1080/07474946.2019.1648922>.
- BARRETO-SOUZA, W. Mixed Poisson INAR(1) processes. *Statistical papers*, [London], v. 60, n. 6, p. 2119–2139, 2017. DOI: <https://doi.org/10.1007/s00362-017-0912-x>.
- BENJAMIN, M.; RIGBY, R.; STASINOPOULOS, M. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, Washington, v. 98, n. 461, p. 214-223, 2003. DOI: <https://doi.org/10.1198/016214503388619238>.
- BOX, G.; PIERCE, D. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, Washington, v. 65, n. 332, p. 1509-1526, 1970. DOI: <https://doi.org/10.2307/2284333>.
- BRASIL. Ministério da Saúde. *Mortalidade - Brasil*. Brasília: MS, [2022]. Available from: <http://tabnet.datasus.gov.br/>. Access in: Jan. 2022.
- BROEK, J. A score test for zero inflation in a Poisson distribution. *Biometrics*, New York, n. 1, p. 738-743, 1995. DOI: <https://doi.org/10.2307/2532959>.
- BROEMELING, L. *Bayesian analysis of time series*. New York: CRC Press, 2019.
- CANOVA, F.; HANSEN, B. Are seasonal patterns constant over time? A test for seasonal stability. *Journal of business & economic statistics*, Washington, v. 13, n. 3, p. 237–252, 1995. DOI: <https://doi.org/10.2307/1392184>.

- CHIVERS, C. *MHadaptive*: General Markov chain Monte Carlo for Bayesian inference using adaptive Metropolis-Hastings sampling. 2015. Available from: <https://CRAN.R-project.org/package=MHadaptive>. Access in: Jan. 2022.
- COX, D.; STUART, A. Some quick sign tests for trend in location and dispersion. *Biometrika*, Oxford, v. 42, n. 1/2, p. 80-95, 1955. DOI: <https://doi.org/10.2307/2333424>.
- DUANE, S.; KENNEDY, A.; PENDLETON, B.; ROWETH, D. Hybrid Monte Carlo. *Physics letters B*, Amsterdam, v. 195, n. 2, p. 216-222, 1987. DOI: [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X).
- FENG, C. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *Journal of statistical distributions and applications*, Heidelberg, v. 8, n. 1, p. 1-19, 2021. DOI: <https://doi.org/10.1186/s40488-021-00121-4>.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Staff report*, [s. l.], n. 148, p. 1-29, 1991.
- GHAHRAMANI, M.; WHITE, S. Time series regression for zero-inflated and overdispersed count data: A functional response model approach. *Journal of statistical theory and practice*, Greensboro, v. 14, n. 2, p. 1-18, 2020. DOI: <https://doi.org/10.1007/s42519-020-00094-8>.
- GONÇALVES, J.; BARRETO-SOUZA, W. Flexible regression models for counts with high-inflation of zeros. *Metron*, [s. l.], v. 78, n. 1, p. 71-95, 2020. DOI: <https://doi.org/10.1007/s40300-020-00163-9>.
- HASHIM, L.; HASHIM, K.; SHIKER, M. An application comparison of two Poisson models on zero count data. *Journal of physics*, [Bristol], v. 1818, n. 1, 2021. DOI: <https://doi.org/10.1088/1742-6596/1818/1/012165>.
- HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. *Operations research*, Switzerland, v. 31, n. 6, p. 1109-1144, 1983. DOI: [https://doi.org/0030-364X/83/3106-1109\\$01.25](https://doi.org/0030-364X/83/3106-1109$01.25).
- HILBE, J. *Modeling count data*. Cambridge University Press, 2014.
- KORENEV, B. *Bessel functions and their applications*. New York: CRC Press, 2002.
- LAMBERT, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, Richmond v. 34, n. 1, p. 1-14, 1992. DOI: <https://doi.org/10.2307/1269547>.
- LÜDECKE, D.; BEN-SHACHAR, M.; PATIL, I.; WAGGONER, P.; MAKOWSKI, D. An R package for assessment, comparison and testing of statistical models. *Journal of open source software*, Chicago, v. 6, n. 60, p. 31-39, 2021. DOI: <https://doi.org/10.21105/joss.03139>.
- MULLAHY, J. Specification and testing of some modified count data models. *Journal of econometrics*, Amsterdam, v. 33, n. 3, p. 341-365, 1986. DOI: [https://doi.org/10.1016/0304-4076\(86\)90002-3](https://doi.org/10.1016/0304-4076(86)90002-3).
- NELDER, J.; WEDDERBURN, R. Generalized linear models. *Journal of the royal statistical society*, London, v. 135, p. 370-384, 1972. DOI: <https://doi.org/10.2307/2344614>.
- PAYNE, E.; HARDIN, J.; EGEDE, L.; RAMAKRISHNAN, V.; SELASSIE, A.; GEBREGZIABHER, M. Approaches for dealing with various sources of over dispersion in modeling count data: Scale adjustment versus modeling. *Statistical methods in medical research*, Singapore v. 26, n. 4, p. 1802-1823, 2017. DOI: <https://doi.org/10.1177/0962280215588569>.
- RAFTERY, A.; LEWIS, S. Comment: One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical science*, [Commack], v. 7, n. 4, p. 493-497, 1992. DOI: <https://doi.org/10.1214/ss/1177011143>.
- RIGBY, R.; STASINOPOULOS, M.; HELLER, G.; DE BASTIANI, F. *Distributions for modeling location, scale, and shape*. New York: Chapman and Hall: CRC, 2019.
- ROBERT, C. Modified Bessel functions and their applications in probability and statistics. *Statistics & probability letters*, Amsterdam, v. 9, n. 2, p. 155-161, 1990. DOI: [https://doi.org/10.1016/0167-7152\(92\)90011-S](https://doi.org/10.1016/0167-7152(92)90011-S).
- ROCHA, A.; CRIBARI-NETO, F. Beta autoregressive moving average models. *Test*, [London], v. 18, n. 3, p. 529-545, 2009. DOI: <https://doi.org/10.1007/s11749-008-0112-z>.

- SÁFADI, T.; MORETTIN, P. A Bayesian analysis of autoregressive models with random normal coefficients. *Journal of statistical computation and simulation*, Philadelphia, v. 73, n. 8, p. 563-573, 2003. DOI: <https://doi.org/10.1080/0094965031000136003>.
- SATHISH, V.; MUKHOPADHYAY, S.; TIWARI, R. Autoregressive and moving average models for zero-inflated count time series. *Statistica Neerlandica*, Gravenhage, v. 76, n. 2, p. 190-218, 2021. DOI: <https://doi.org/10.1111/stan.12255>.
- STASINOPOULOS, M.; RIGBY, R. *Gamlss.dist: Distributions for generalized additive models for location scale and shape*. Available from: <https://cran.r-project.org/web/packages/gamlss.dist/gamlss.dist.pdf>. Access in: Jan. 2022.
- TAWIAH, K.; IDDRISU, W.; ASOSEGA, K. Zero-inflated time series modelling of COVID-19 deaths in Ghana. *Journal of environmental and public health*, New York, v. 2021, p. 1-9, 2021. DOI: <https://doi.org/10.1155/2021/5543977>.
- WORLD HEALTH ORGANIZATION. *ICD-10: International statistical classification of diseases and related health problems*. 2. nd. Geneva: WHO, 2004. Available from: <https://apps.who.int/iris/handle/10665/42980>. Access in: Nov. 2022.
- YANG, M.; ZAMBA, G.; CAVANAUGH, J. Markov regression models for count time series with excess zeros: A partial likelihood approach. *Statistical methodology*, [London], v. 14, p. 26-38, 2013. DOI: <https://doi.org/10.1016/j.stamet.2013.02.001>.
- ZUO, G.; FU, K.; DAI, X.; ZHANG, L. Generalized Poisson hurdle model for count data and its application in ear disease. *Entropy*, Basel, v. 23, n. 9, p. 1-16, 2021. DOI: <https://doi.org/10.3390/e23091206>.
- ZUUR, A.; IENO, E.; WALKER, N.; SAVELIEV, A.; SMITH, G. *Mixed effects models and extensions in ecology with R*. New York: Springer, 2009.

Received: Oct. 17, 2022  
 Accepted: Dec. 13, 2022  
 Published: Dec. 19, 2022





***ARTIGO 3 - Distribuição Poisson zero-ajustada com parâmetros variando no tempo para a análise de séries temporais de contagem***

Redigido conforme as normas do *Journal of applied statistics* (versão em processo de submissão).

## Distribuição Poisson zero-ajustada com parâmetros variando no tempo para a análise de séries temporais de contagem

Luiz Otávio de Oliveira Pala <sup>a\*</sup> e Thelma Sáfiadi<sup>a</sup>

<sup>a</sup> Departamento de Estatística. Universidade Federal de Lavras, Minas Gerais, Brasil.

### Resumo

Diversos estudos têm utilizado as extensões dos modelos ARMA(p, q) para a análise de séries temporais não Gaussianas. Uma delas corresponde a *Generalized Autoregressive Moving Average*, GARMA(p, q), possibilitando a modelagem de séries de contagem a partir de distribuições como a Poisson. Na literatura, a classe GARMA(p, q) está sendo ampliada para outras distribuições, com o intuito de comportar as características típicas de contagens, envolvendo sub ou superdispersão e excesso de zeros. Este trabalho tem como objetivo propor uma abordagem baseada na classe GARMA(p, q) para a análise de séries de contagem com excesso de zeros, assumindo distribuição Poisson zero-ajustada com parâmetros variando no tempo, de modo a comportar a correlação serial e permitir realizar previsões de contagens e da probabilidade de zeros. Para a inferência, adotou-se a análise Bayesiana com o uso do algoritmo Monte Carlo Hamiltoniano para a amostragem da posteriori conjunta. Ao longo do estudo, foi realizado um estudo de simulação e uma aplicação em dados de mortalidade em decorrência da influenza. Os resultados da aplicação indicaram a utilidade do modelo ao se estimar a probabilidade de não ocorrência e o número de óbitos em períodos futuros.

### Palavras-chave

Excesso de zeros. GARMA(p, q). Inferência Bayesiana. Influenza.

## 1. Introdução

A análise e a previsão de séries temporais são tópicos ativos na esfera da ciência e de áreas como a engenharia [24]. Em estudos aplicados, os modelos de séries temporais utilizados comumente baseiam-se na classe ARMA(p, q), fundamentada em [6], e suas respectivas extensões, sendo considerados mais apropriados para a modelagem de séries temporais Gaussianas, mas que permitem a realização de previsões satisfatórias, que é uma tarefa de grande importância [24, 35].

Em relação a séries temporais de contagem e de valores discretos, [14] aborda que modelos Gaussianos podem ter o desempenho reduzido ao descrever o comportamento do processo. E isso pode ser uma consequência das características estilizadas destes tipos de dados, pois podem apresentar sub ou superdispersão, ou seja, variância menor ou maior que a média [4, 33]; excesso de zeros [14, 21] e presença de valores atípicos [28], que podem conduzir ao problema de superdispersão [4, 28]; e frequentemente, autocorrelações não negativas [14].

Inicialmente, as séries de contagem eram analisadas a partir da classe tradicional de Modelos Lineares Generalizados [14], sendo esta uma teoria que trouxe contribuições para a fundamentação de metodologias que passaram a comportar a estrutura de de-

pendência temporal. Um exemplo disso é a classe de modelos *Generalized Autoregressive Moving Average*, GARMA(p, q), criada por [5] no ano de 2003, que introduziu uma estrutura de dependência na forma ARMA(p, q) no preditor linear em distribuições pertencentes à família exponencial como a Poisson e a Binomial negativa.

A classe GARMA(p, q) tem sido utilizada e ampliada em diversos estudos, abrangendo perspectivas via inferência Bayesiana [2], extensões para séries que apresentem comportamentos sazonais [8] e flexibilização para distribuições como a *Conway-Maxwell-Poisson* (CMP) e a *Bernoulli-geometric* (BerG), que são utilizadas para comportar situações de variações na dispersão dos dados [14, 18, 26, 33].

No contexto de séries temporais de contagem com excesso de zeros, as distribuições tradicionais, como a Poisson, a Binomial e a Binomial negativa, podem não comportar tal situação [14]. Elevadas proporções de zero não podem ser desprezadas durante a modelagem pois podem prejudicar a inferência e conduzir a relações espúrias [1]. Neste sentido, estudos têm explorado distribuições infladas de zero, como Poisson (ZIP) e a Binomial negativa (ZINB) infladas de zero, e as suas versões ajustadas em zero como a Poisson zero-ajustada (ZAP) [1, 21, 33, 34, 39].

A ZAP é uma distribuição de contagem que pode ser vista como a mistura de dois componentes, um deles associado com a massa de probabilidade no ponto zero e o outro relacionado aos valores positivos, que segue uma distribuição Poisson truncada em zero [19]. Diversas aplicações com a distribuição ZAP podem ser encontradas na literatura tal como em [3, 19, 22, 33], além de ampliações para a modelagem de contagens multivariadas [40].

Este trabalho tem como objetivo propor uma abordagem baseada na classe GARMA(p, q) para a modelagem de séries temporais de contagem que apresentem excesso de zeros, assumindo que a variável resposta segue uma distribuição ZAP em que os parâmetros variam ao longo do tempo de modo a comportar a correlação serial. Para o processo inferencial, adotou-se uma análise Bayesiana similar àquela utilizada por [2], no entanto, recorrendo ao algoritmo Monte Carlo Hamiltoniano (HMC), proposto em [16], para a amostragem da posteriori conjunta.

A organização deste trabalho é dada da seguinte forma: na Seção (2) tem-se a definição do modelo e a análise Bayesiana utilizada está descrita na Seção (3). Um estudo de simulação é apresentado na Seção (4) e uma aplicação a dados de mortalidade em decorrência da influenza é desenvolvida na Seção (5). Por fim, as considerações finais estão apresentadas na Seção (6).

## 2. Definição do modelo

Denote  $Y$  como uma série temporal igualmente espaçada e indexada no tempo  $t = \{1, \dots, n\}$  e  $\mathcal{F}_{t-1} = \{y_1, y_2, \dots, y_{t-1}, \mu_1, \mu_2, \dots, \mu_{t-1}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{t-1}\}$  o conjunto de informações passadas até o instante  $t - 1$ , em que  $\mathbf{x}_t = (\mathbf{x}_{t1}, \dots, \mathbf{x}_{tr})^\top$  é o vetor contendo  $r$  variáveis explicativas que estão associadas com o vetor de coeficientes  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_r)^\top$ .

Defina  $\Phi_p(B) = (1 - \phi_1 B^1 - \dots - \phi_p B^p)$  e  $\Theta_q(B) = (1 - \theta_1 B^1 - \dots - \theta_q B^q)$  os polinômios autorregressivo e de médias móveis de ordens  $p$  e  $q$ , respectivamente. De forma similar, simbolize  $\Phi_P(B^s) = (1 - \Phi_1 B^{s1} - \dots - \Phi_P B^{sP})$  e  $\Theta_Q(B^s) = (1 - \Theta_1 B^{s1} - \dots - \Theta_Q B^{sQ})$  como o polinômio autorregressivo e o de médias móveis sazonal de ordens  $P$  e  $Q$ , respectivamente. Nessa notação  $B$  representa o operador de retardo e  $s$  é o tamanho do período.

Assumindo que  $y_t \mid \mathcal{F}_{t-1}$  segue uma distribuição Poisson zero-ajustada, ZAP( $\mu_t$ ,

$\nu_t$ ), a densidade condicional pode ser expressa conforme mostrado na Equação (1).

$$p(y_t | \mathcal{F}_{t-1}) = \nu_t \mathbb{I}_{(y_t=0)} + \left[ \frac{1 - \nu_t}{1 - e^{-\mu_t}} \right] \left[ \frac{e^{-\mu_t} \mu_t^{y_t}}{y_t!} \right] \mathbb{I}_{(y_t>0)}, \quad y_t \in \mathbb{N}, \quad (1)$$

definida em  $\Omega = \{\mu_t, \nu_t \mid \mu_t > 0, 0 < \nu_t < 1\}$ , sendo  $\mathbb{I}_{(\cdot)}$  a função indicadora,  $\nu_t$  a probabilidade de que a contagem seja zero no instante  $t$  e  $\mu_t$  a esperança condicional na situação em que  $y_t > 0$  [30]. Com o condicionamento de  $y_t$  à  $\mathcal{F}_{t-1}$ , a esperança é dada por  $\frac{(1-\nu_t)\mu_t}{1-e^{-\mu_t}}$ , sendo uma ponderação de  $\mu_t$  por  $\frac{1-\nu_t}{1-e^{-\mu_t}}$ , que é a razão da probabilidade exata complementar de que  $y_t = 0$  e a probabilidade de que  $y_t > 0$ , proveniente de uma distribuição Poisson com média  $\mu_t$  [30].

Utilizando a função de ligação logarítmica, o preditor linear baseado na classe GARMA(p, q) pode ser escrito conforme a Equação (2):

$$\begin{aligned} \log(\mu_t) = & \Phi_p(B)\Phi_p(B^s) \left[ \mathbf{x}_t^\top \boldsymbol{\beta} - \log(y_t) \right] \\ & - \Theta_q(B)\Theta_q(B^s) [\log(y_t) - \log(\mu_t)] + \log\left(\frac{y_t^2}{\mu_t}\right), \end{aligned} \quad (2)$$

que pode ser generalizado com a inserção do operador de integração, resultando no processo SARIMA(p, d, q)(P, D, Q)<sub>s</sub> [8]. Veja que pode ser necessário substituir  $y_{t-j}$  por  $y_{t-j}^* = \max(y_{t-j}; 0.10)$ , garantindo a existência da função de ligação no ponto zero.

Ademais, considere a existência de uma relação entre  $\nu_t$  e as defasagens da série temporal, segundo o processo  $\text{logit}(\nu_t) = \omega_0 + \sum_{j=0}^J \omega_j y_{t-j}$ , em que  $\boldsymbol{\omega} = (\omega_0, \dots, \omega_J)^\top$  é o vetor de coeficientes associado à  $\nu_t$ . Essa especificação faz com que o parâmetro associado com a probabilidade de contagens iguais a zero,  $\nu_t$ , varie no tempo e permita a realização de previsões dessa probabilidade em instantes  $t+h$ . A suposição de que  $\nu_t$ , assim como  $\mu_t$ , varie no tempo, pode ser vista como uma ampliação de um dos modelos utilizados por [33], que considerou que o parâmetro  $\nu$  da ZAP é fixo no tempo ao modelar a série do número mensal de chamadas telefônicas.

Denotando como  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\omega})^\top$  o conjunto de todos os parâmetros do modelo, então a função de verossimilhança aproximada e condicionada às  $m = \max(p, q, sP, sQ, J)$  primeiras observações é dada por  $L(\boldsymbol{\theta} | Y) \approx \prod_{t=m+1}^n p(y_t | \mathcal{F}_{t-1})$ , podendo ser necessário supor que os primeiros termos do erro são iguais a zero, assim como descrito nos estudos de [5, 31]. Observe que, a cada passo, a média  $\mu_t$  e a probabilidade  $\nu_t$  podem ser determinadas a partir dos seus respectivos componentes que estão presentes no vetor  $\boldsymbol{\theta}$ .

### 3. Análise Bayesiana

Tratando-se de modelagem, a estimativa de  $\boldsymbol{\theta}$  em modelos GARMA(p, q) pode ser obtida maximizando a função de verossimilhança aproximada a partir de um método numérico. Para esse procedimento, a função *garmaFit* do pacote *gamlss.util* [38], permite a realização da análise, mas sem a inserção de ordens sazonais. Além disso, pode ser adotada a abordagem Bayesiana, que apresentou um bom desempenho em termos de estimativa pontual e intervalar no estudo de [2].

Conforme o Teorema de Bayes, a distribuição a posteriori ( $\boldsymbol{\pi}(\boldsymbol{\theta} | Y)$ ) é proporcional ao produto  $L(\boldsymbol{\theta} | Y) \pi_0(\boldsymbol{\theta})$ , sendo  $L(\boldsymbol{\theta} | Y)$  a função de verossimilhança aproximada condicionada às  $m$  primeiras observações e  $\pi_0(\boldsymbol{\theta})$  a distribuição a priori conjunta de  $\boldsymbol{\theta}$ .

Para essa análise, foram consideradas prioris não informativas e independentes para todos os componentes de  $\boldsymbol{\theta}$ .

Para o conjunto de parâmetros associados às variáveis explicativas, foi estabelecido que cada componente de  $\boldsymbol{\beta}$  é normalmente distribuído, isto é:

$$p(\beta_j) \propto \exp \left[ -\frac{1}{2} \left( \frac{\beta_j - \mu_j}{\tau_j} \right)^2 \right], \quad \beta_j \in (-\infty, \infty), j = \{1, \dots, r\},$$

sendo  $\mu_j = 0$  e  $\tau_j = 10$  os hiperparâmetros associados com a distribuição a priori para cada  $\beta_j$ . A mesma estrutura de prioris adotada em  $\boldsymbol{\beta}$  foi utilizada em  $\boldsymbol{\Phi}$ ,  $\boldsymbol{\Theta}$  e  $\boldsymbol{\omega}$ , ou seja, prioris normais com média zero e alta variabilidade.

A inferência foi realizada via amostragem da posteriori conjunta com o algoritmo HMC, disponível no *RStan* [36]. O HMC utiliza a dinâmica Hamiltoniana para a amostragem, suprindo o comportamento de passeio aleatório local observado em outros algoritmos [20] e recorrendo ao uso da informação do gradiente como forma de contribuir com a descrição da posteriori [12]. [10, 25] destacam que o HMC produz amostras eficientes, tem desempenho satisfatório na situação de elevada dimensão de  $\boldsymbol{\theta}$  e baixa correlação entre as amostras.

A estratégia adotada pelo HMC é provocar o aumento da posteriori inserindo  $\mathbf{p} \sim N_d(\mathbf{0}, \mathbf{M})$  parâmetros auxiliares de modo que  $\boldsymbol{\theta} \in \mathbb{R}^d$ ,  $\mathbf{p} \in \mathbb{R}^d$  com  $\boldsymbol{\theta} \perp \mathbf{p}$ , sendo  $\mathbf{M}$  frequentemente diagonal [20]. O motivo para esse aumento é exclusivamente possibilitar que o algoritmo faça uma exploração mais rápida do espaço dos parâmetros [12, 20]. Portanto, a posteriori aumentada é o produto das marginais de  $\boldsymbol{\theta}$  e  $\mathbf{p}$ , resultando em  $p(\boldsymbol{\theta}, \mathbf{p} | Y) = p(\mathbf{p})\pi(\boldsymbol{\theta} | Y)$ , em que o interesse está apenas em  $\boldsymbol{\theta}$  [20].

Com a caracterização da  $p(\boldsymbol{\theta}, \mathbf{p} | Y)$  constrói-se a função Hamiltoniana dada por  $H(\boldsymbol{\theta}, \mathbf{p}) = -\log(p(\boldsymbol{\theta}, \mathbf{p} | Y)) = -\log(\pi(\boldsymbol{\theta} | Y)) + \frac{1}{2} \log((2\pi)^d | \mathbf{M} |) + \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}$ , cujas derivadas parciais  $\frac{\partial \boldsymbol{\theta}}{\partial t} = \frac{\partial H(\boldsymbol{\theta}, \mathbf{p})}{\partial \mathbf{p}}$  e  $\frac{\partial \mathbf{p}}{\partial t} = -\frac{\partial H(\boldsymbol{\theta}, \mathbf{p})}{\partial \boldsymbol{\theta}}$ , chamadas de equações Hamiltonianas, são responsáveis por mapear o estado do processo do instante  $t$  para  $t + z$  e são solucionadas via método de Störmer-Verlet (*leapfrog*) em  $L$  passos, de modo a propor o novo estado da cadeia [12, 27]. O estado candidato é avaliado via regra aceitação/rejeição de Metropolis [20, 27]. Assim como em outros métodos Monte Carlo via cadeias de Markov, esse procedimento é repetido até a convergência para a distribuição estacionária [20].

A convergência para a distribuição estacionária pode ser avaliada através de algum procedimento. Um dos métodos disponíveis no pacote *RStan* [36] é o  $\hat{R}$ -split, proposto por [20], que pode ser visto como uma ampliação do Fator de Redução de Escala Potencial ( $\hat{R}$ ). O procedimento requer a construção de múltiplas cadeias iniciadas arbitrariamente para tomar a decisão acerca da convergência. Conforme o diagnóstico, o processo pode ser considerado convergente para valores em que  $\hat{R}$ -split  $< 1,05$  [36].

Todavia, a amostragem de algumas distribuições pode ser complicada para qualquer algoritmo [25]. No caso do HMC, a matriz  $\mathbf{M}$  e o número de passos  $L$  podem ser calibrados durante o período de aquecimento, o que pode resultar em um melhor desempenho [12, 23]. Conforme [20], uma das formas de se realizar a calibração de  $\mathbf{M}$  é adaptá-la utilizando o inverso da matriz de covariâncias da posteriori. Ademais, uma extensão ao HMC, que usa uma forma recursiva para adaptar  $L$  em cada iteração, foi proposta por [23] e denominada como *No-U-Turn Sampler* (NUTS), cujos detalhes estão em [23, 36].

### 3.1. Realização de previsões

Para realizar previsões de  $Y$  em um horizonte  $h$  é necessário definir a distribuição preditiva, isto é, a distribuição de  $y_{t+h}$  condicionada a todos os parâmetros e observações passadas do processo [9, 32]. Combinando a posteriori conjunta  $\pi(\boldsymbol{\theta} | Y)$  com a densidade da nova observação,  $p(y_{t+h} | \boldsymbol{\theta}, \mathcal{F}_{t+h-1})$ , a densidade preditiva é dada por:

$$p(y_{t+h} | \mathcal{F}_{t+h-1}) = \int_{\boldsymbol{\theta} \in \Omega} p(y_{t+h} | \boldsymbol{\theta}, \mathcal{F}_{t+h-1}) \pi(\boldsymbol{\theta} | Y) d\boldsymbol{\theta},$$

que não apresenta forma fechada. No entanto, pode-se obter uma aproximação Monte Carlo da preditiva a partir de  $\mathcal{N}$  amostras de  $\boldsymbol{\theta}^i$ ,  $i = \{1, \dots, \mathcal{N}\}$ , da seguinte forma:

$$p(y_{t+h} | \mathcal{F}_{t+h-1}) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} p(y_{t+h} | \boldsymbol{\theta}^i, \mathcal{F}_{t+h-1}),$$

cujo o valor esperado de  $y_{t+h}$  é dado por:

$$E(y_{t+h}) = \int_{y_{t+h} \in \Omega} y_{t+h} p(y_{t+h} | \mathcal{F}_{t+h-1}) dy_{t+h}. \quad (3)$$

Para o modelo ZAP( $\mu_t, \nu_t$ ), a aproximação da Expressão (3) foi obtida a partir da esperança condicional  $E(y_t | \mathcal{F}_{t-1})$ , isto é:

$$\widehat{y}_{t+h} \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \frac{(1 - \theta_4^i) \mu_{t+h}(\theta_1^i, \theta_2^i, \theta_3^i, \mathcal{F}_{t+h-1})}{1 - e^{-\mu_{t+h}(\theta_1^i, \theta_2^i, \theta_3^i, \mathcal{F}_{t+h-1})}},$$

sendo  $\boldsymbol{\theta} = (\beta, \Phi, \Theta, \omega)^\top$ . Esse resultado corresponde a média de  $p(y_{t+h} | \mathcal{F}_{t+h-1})$  atualizada para cada  $\boldsymbol{\theta}^i$ .

Além disso, a probabilidade futura de não ocorrência de contagens iguais a zero,  $\widehat{\nu}_{t+h}$ , pode ser aproximada de forma análoga àquela utilizada para a  $E(y_{t+h})$ , isto é:

$$\widehat{\nu}_{t+h} \approx \sum_{i=1}^{\mathcal{N}} \frac{\exp\left(\omega_0^i + \sum_{j=0}^J \omega_j^i y_{t-j}\right)}{1 + \exp\left(\omega_0^i + \sum_{j=0}^J \omega_j^i y_{t-j}\right)},$$

em que  $\widehat{\nu}_{t+h} \in (0, 1)$  e representa a probabilidade média a posteriori de que a série assumo valor zero em um instante  $t + h$ .

## 4. Um estudo de simulação

Para o estudo de simulação, foram considerados os modelos e as configurações de parâmetros apresentados na Tabela 1. Observe que os cenários I e II compreendem modelos com estrutura de dependência autorregressiva de ordem dois, enquanto as configurações III e IV supõem que a estrutura de dependência da média é dada por um processo autorregressivo sazonal de primeira ordem.

Os valores de  $\phi_1$ ,  $\phi_2$  e  $\Phi_1$  foram selecionados de modo a manter as condições de estacionariedade dos processos na média, ou seja, situações em que as raízes do polinômio característico estão fora do círculo unitário. Adotou-se que  $\nu_t$  depende do instante imediatamente anterior da série, ou seja,  $\text{logit}(\nu_t) = \omega_0 + \omega_1 y_{t-1}$ . Os parâmetros  $\omega_0$  e  $\omega_1$  foram fixados de modo a possibilitar variações da proporção de zeros. Por exemplo, caso  $y_{t-1} = 0$  no cenário II, a probabilidade da contagem ser zero no instante  $t$  é de aproximadamente 11,92%.

**Tabela 1.** Configurações de parâmetros para o modelo  $\text{ZAP}(\mu_t, \nu_t)$  utilizados no estudo de simulação.

Configuração	n	$\omega_0$	$\omega_1$	$\phi_1$	$\phi_2$	$\Phi_1$
I	{150; 200}	-2,00	0,60	0,50	-0,30	-
II	{150; 200}	-2,00	-1,00	0,80	-0,40	-
III	{150; 200}	-2,00	0,60	0,50	-	0,50
IV	{150; 200}	-1,00	-0,50	-0,80	-	0,20

Fonte: Autores (2022).

A simulação foi realizada no programa R [29] utilizando o pacote *gamlss.dist* [37] para a geração das séries pseudo-aleatórias e o *RStan* [36] para a amostragem da posteriori conjunta via algoritmo HMC. Cada modelo foi replicado  $\omega = 1.000$  vezes e para a amostragem foram estipuladas 10.000 amostras, período *warm-up* de 4.000, *thin* igual a um e duas cadeias. Foram adotados  $L = 10$  passos e matriz  $\mathbf{M}$  adaptada durante o período *warm-up*, que são parâmetros padrão do *RStan*. A convergência foi verificada em cada réplica via diagnóstico  $\hat{R}$ -split, admitindo como convergentes os processos em que  $\hat{R}$ -split  $< 1,05$ .

Foram armazenadas as estimativas de média, moda e desvio padrão a posteriori dos parâmetros de cada modelo replicado, além das estatísticas *corrected error* (CE) e *corrected bias* (CB), que permitem avaliar o desempenho inferencial, assim como realizado no estudo de [2]. As estatísticas CE e CB foram calculadas da seguinte forma  $CE^2 = \frac{1}{w\tau^2} \sum_{i=1}^w (\hat{\theta}^i - \theta)^2$  e  $CB = \frac{1}{w} \sum_{i=1}^w \left| \frac{\theta - \hat{\theta}^i}{\theta} \right|$ , sendo  $\tau$  o desvio padrão amostral de  $\theta$  entre as  $\omega$  réplicas. Espera-se que, no processo de simulação, as estimativas de CE e CB se aproximem de um e de zero, respectivamente.

Os resultados de simulação das configurações I e II estão apresentados na Tabela 2. Pode-se observar uma redução dos valores da estatística CB e dos desvios padrão a posteriori a medida em que o tamanho amostral passou de 150 para 200 observações. Analogamente, os valores da estatística CE aproximaram-se de um com o aumento do tamanho amostral, indicando uma melhor aproximação dos valores estimados.

Ainda em relação aos resultados da Tabela 2, observa-se que as estimativas da moda a posteriori obtiveram um melhor desempenho quando comparadas com as da média na maioria dos cenários avaliados. A única situação em que a estimativa da média apresentou melhor performance que a da moda foi ao avaliar o parâmetro  $\omega_0$  na configuração II com  $n = 200$ , que encontra-se destacada na Tabela 2.

Ao avaliar os resultados de simulação das configurações III e IV, correspondendo a modelos com componentes sazonais, pode-se observar por meio da Tabela 3 que os resultados da estatística CB concordaram com os resultados das configurações I e II, isto é, apresentando redução à medida em que o tamanho amostral cresceu. Os valores da estatística CE aproximaram-se de um, mas apresentando aumento em duas situações associadas ao parâmetro  $\phi_1$ , cujas situações estão destacadas na Tabela 3.

Pode-se verificar na Tabela 3 que não houve grande superioridade do desempenho das estimativas da moda a posteriori quando comparadas com as da média nos cenários

**Tabela 2.** Resultados da simulação dos modelos ZAP( $\mu_t, \nu_t$ ) com estrutura AR(2) em  $\mu_t$ , relativo aos cenários I e II, considerando  $w = 1.000$  réplicas de cada modelo.

Configuração	n	$\theta$	Real	Média	Moda	DP	CB	CE
I	150	$\omega_0$	-2,000	-2,082	-2,055	0,351	0,143	1,026
		$\omega_1$	0,600	0,644	0,634	0,184	0,265	1,024
		$\phi_1$	0,500	0,511	0,507	0,129	0,190	1,004
		$\phi_2$	-0,300	-0,290	-0,295	0,065	0,176	1,011
	200	$\omega_0$	-2,000	-2,042	-2,023	0,299	0,121	1,009
		$\omega_1$	0,600	0,620	0,612	0,156	0,219	1,007
		$\phi_1$	0,500	0,508	0,504	0,111	0,171	1,002
		$\phi_2$	-0,300	-0,291	-0,295	0,056	0,150	1,011
II	150	$\omega_0$	-2,000	-2,070	-1,992	1,102	0,339	1,002
		$\omega_1$	-1,000	-1,448	-1,242	0,884	0,731	1,075
		$\phi_1$	0,800	0,767	0,774	0,110	0,111	1,045
		$\phi_2$	-0,400	-0,389	-0,396	0,112	0,213	1,004
	200	$\omega_0$	-2,000	<b>-2,016</b>	<b>-1,961</b>	0,887	0,291	1,000
		$\omega_1$	-1,000	-1,308	-1,167	0,710	0,552	1,077
		$\phi_1$	0,800	0,777	0,782	0,094	0,096	1,029
		$\phi_2$	-0,400	-0,394	-0,400	0,096	0,199	1,001

Fonte: Autores (2022).

III e IV. Em relação aos desvios padrão, os resultados indicaram redução da magnitude dos desvios conforme o tamanho amostral cresceu, o que também ocorreu nas configurações I e II.

De um modo geral, os resultados dos cenários avaliados computacionalmente indicaram boas propriedades da inferência em termos da redução dos valores de CB conforme o tamanho amostral cresceu, mas evidenciaram certa alternância entre o desempenho dos estimadores da média e moda a posteriori principalmente nas configurações III e IV.

## 5. Análise de dados de mortalidade em decorrência da influenza

Para a aplicação foram utilizados os dados do número de óbitos, de internações e da média do tempo de internação de idosos com idade igual ou superior a 70 anos, em decorrência da influenza, no município de São Paulo, Brasil. Os dados foram disponibilizados pelo Departamento de Informática do Sistema Único de Saúde e abrangem o período entre janeiro de 2008 e setembro de 2022 [15].

No período foram reportados 97 óbitos, resultando em uma média mensal de 0,548 óbitos em decorrência da influenza. No entanto, como pode ser observado na Figura 1(a), contagens iguais a zero ocorreram em 75,10% dos meses, caracterizando uma situação de excesso de zeros. Ao desconsiderar os meses com ausência de óbitos, a média mensal é de aproximadamente dois óbitos.

O comportamento da função de autocorrelação (ACF) da série temporal do número de óbitos, disponível na Figura 1(b), sugere uma série não estacionária. A hipótese nula de ausência de tendência foi não rejeitada pelo teste de Cox Stuart (valor-p = 0,405) [13]. De modo análogo, o teste de Canova e Hansen [11] não sugeriu a presença da componente sazonal (valor-p = 0,402).

O comportamento das séries explicativas que foram utilizadas para modelar o

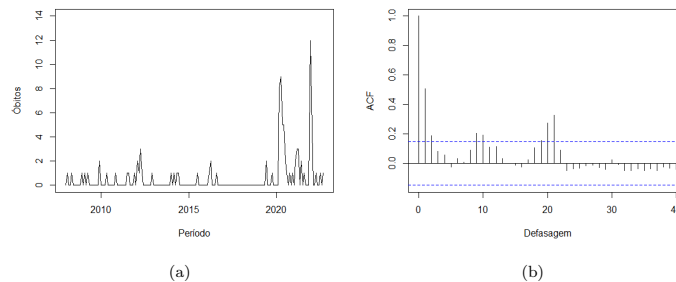


**Tabela 3.** Resultados da simulação dos modelos ZAP( $\mu_t, \nu_t$ ) com estrutura SAR(2), relativo aos cenários III e IV, considerando  $w = 1.000$  réplicas de cada modelo.

Configuração	n	$\theta$	Real	Média	Moda	DP	CB	CE
III	150	$\omega_0$	-2,000	-2,115	-2,085	0,365	0,153	1,044
		$\omega_1$	0,600	0,654	0,641	0,192	0,277	1,034
		$\phi_1$	0,500	0,495	0,485	0,116	0,176	<b>1,000</b>
		$\Phi_1$	0,500	0,486	0,478	0,108	0,165	1,009
	200	$\omega_0$	-2,000	-2,063	-2,043	0,306	0,129	1,018
		$\omega_1$	0,600	0,632	0,622	0,162	0,230	1,016
		$\phi_1$	0,500	0,493	0,485	0,100	0,151	<b>1,002</b>
		$\Phi_1$	0,500	0,489	0,483	0,093	0,141	1,007
IV	150	$\omega_0$	-1,000	-0,949	-0,948	0,360	0,275	1,010
		$\omega_1$	-0,500	-0,612	-0,558	0,267	0,422	1,092
		$\phi_1$	-0,800	-0,794	-0,795	0,045	0,045	<b>1,009</b>
		$\Phi_1$	0,200	0,201	0,196	0,074	0,282	1,000
	200	$\omega_0$	-1,000	-0,988	-0,988	0,311	0,241	1,000
		$\omega_1$	-0,500	-0,578	-0,536	0,228	0,348	1,060
		$\phi_1$	-0,800	-0,794	-0,795	0,038	0,039	<b>1,013</b>
		$\Phi_1$	0,200	0,201	0,198	0,063	0,250	1,000

Fonte: Autores (2022).

**Figura 1.** Comportamento do número de óbitos de idosos com idade igual ou superior a 70 anos, em decorrência da influenza, no município de São Paulo, Brasil.

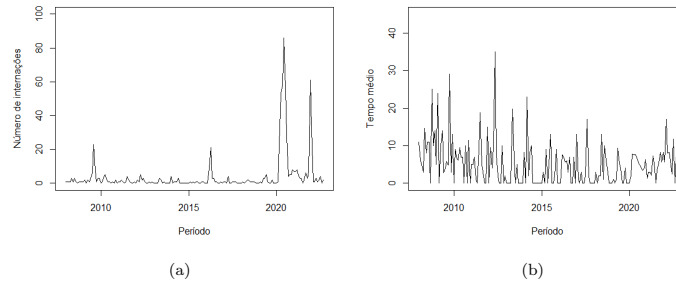


Fonte: Autores (2022).

número de óbitos está disposto na Figura 2. Em 2(a) tem-se o número de internações mensais, apresentando média de aproximadamente 4 internações por mês, e em 2(b) tem-se a média mensal do tempo de internação. O tempo médio de hospitalização foi estimado em 5,077 dias, com máximo de 35 dias. Conforme os dados disponíveis, essas internações resultaram em um custo médio de R\$ 1.111,57 e um custo total de R\$ 772.540,90 para o sistema de saúde.

Para a análise do número de óbitos ( $y_t$ ) foi considerada a primeira defasagem do número de internações ( $x_{t1}$ ) e da média do tempo de hospitalização ( $x_{t2}$ ) como variáveis explicativas, dispostas em  $\mathbf{x}_t$ , além de uma estrutura autorregressiva de segunda ordem.

**Figura 2.** Comportamento do número de internações (a) e do tempo médio de hospitalização (b) de idosos com idade igual ou superior a 70 anos, em decorrência da influenza, no município de São Paulo, Brasil.



Fonte: Autores (2022).

Isto é, supõem-se que  $y_t | \mathcal{F}_{t-1} \sim ZAP(\mu_t, \nu_t)$ , em que:

$$\begin{aligned} \log(\mu_t) &= (1 - \phi_1 B - \phi_2 B^2) \left[ \mathbf{x}_{t-1}^\top \boldsymbol{\beta} - \log(y_t) \right] + \log(y_t), \\ \text{logit}(\nu_t) &= \omega_0 + \omega_1 y_{t-1}. \end{aligned} \quad (4)$$

A amostragem do modelo (4) foi realizada via HMC considerando 10.000 iterações,  $warm-up = 4.000$ ,  $thin = 2$  e 3 cadeias. Conforme os resultados, o parâmetro  $\beta_2$ , associado com o tempo médio de hospitalização, não foi estatisticamente significativo ao nível de 95% de credibilidade, sugerindo que maiores tempos médios de internação não associaram-se com o aumento de óbitos. Desta forma, a variável tempo de hospitalização foi removida do modelo e os respectivos resultados encontram-se na Tabela 4.

A estimativa de  $\beta_1$  sugere uma associação positiva entre o número de internações em decorrência da influenza no período  $t - 1$  e do número de óbitos no mês  $t$ , sendo estatisticamente significativo. Além do parâmetro  $\beta_1$ , os parâmetros associados com a estrutura de dependência que compõem  $\mu_t$ , isto é,  $\phi_1$  e  $\phi_2$ , também foram significativos ao mesmo nível de credibilidade.

**Tabela 4.** Parâmetros estimados para o modelo  $ZAP(\mu_t, \nu_t)$  ajustado a série temporal do número de óbitos em decorrência da influenza.

Modelo	Parâmetro	Média	SD	HPD (95%)		$\hat{R}$ -split
				$L_i$	$L_s$	
ZAP( $\mu_t, \nu_t$ )	$\beta_1$	0,018	0,004	0,009	0,025	1,000
	$\phi_1$	0,187	0,082	0,030	0,353	1,000
	$\phi_2$	-0,266	0,072	-0,406	-0,119	1,000
	$\omega_0$	1,376	0,203	0,990	1,782	1,000
	$\omega_1$	-0,427	0,149	-0,751	-0,166	1,000

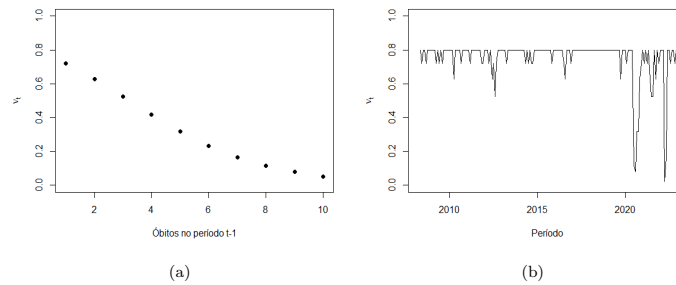
Fonte: Autores (2022).

As estimativas de  $\omega_0$  e  $\omega_1$  indicam que caso não ocorram óbitos em um determinado mês  $t - 1$ , a probabilidade de não ocorrência de nenhum óbito no mês  $t$  é de 0,798. Pode-se verificar uma redução desta probabilidade a medida em que o número de óbitos em  $t - 1$  aumenta. Por exemplo, caso ocorram 5 óbitos em um determinado mês

$t - 1$ , a estimativa da probabilidade de não ocorrência de óbitos no mês  $t$  é de 0,318.

A Figura 3(a) apresenta a probabilidade de que o número de óbitos seja zero no mês  $t$  conforme o número de óbitos reportado no mês  $t - 1$ , estimada a partir da equação  $\text{logit}(\nu_t) = \hat{\omega}_0 + \hat{\omega}_1 y_{t-1}$ . O comportamento latente de  $\nu_t$  ao longo do período estudado está apresentado na Figura 3(b), sendo possível verificar períodos de redução da probabilidade de não ocorrência de óbitos, como em maio de 2020 e janeiro de 2022, meses no qual  $\hat{\nu}_t$  foi estimado em 0,078 e 0,023, respectivamente.

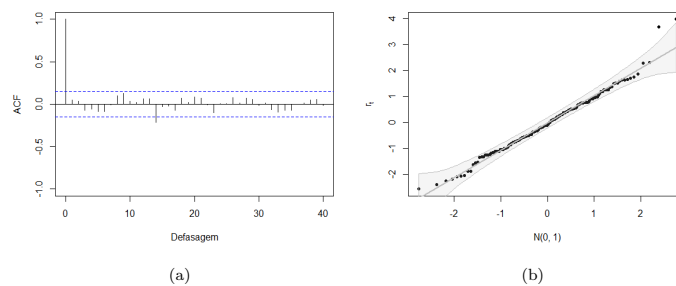
**Figura 3.** Efeito estimado de  $\nu_t$  em relação do número de óbitos em um determinado mês  $t - 1$  (a) e efeito latente de  $\nu_t$  ao longo do período estudado.



Fonte: Autores (2022).

Ao avaliar o comportamento dos resíduos quantílicos randomizados normalizados ( $r_t$ ), desenvolvidos por [17], pode-se verificar, por meio da Figura 4, que estes apresentaram comportamento estacionário, como também apontado pelo teste Box Pierce [7] avaliado até a 24 defasagem (valor-p = 0,402). Além disso, a hipótese nula de normalidade dos resíduos foi não rejeitada conforme o teste de Kolmogorov Smirnov (valor-p = 0,776).

**Figura 4.** Função de autocorrelação (a) e gráfico quantil-quantil (b) dos resíduos do modelo ajustado aos dados de óbitos em decorrência de influenza.



Fonte: Autores (2022).

Com a verificação do comportamento estacionário dos resíduos, pode-se realizar previsões do número de óbitos de idosos em decorrência da influenza através do valor

esperado da preditiva, estimado via amostragem Monte Carlo, cujos parâmetros são dados por:

$$\begin{aligned}\log(\mu_t) &= (1 - \phi_1^i B - \phi_2^i B^2)[\beta_1^i x_{t-1,1} - \log(y_t)] + \log(y_t) \\ &= \beta_1^i x_{t-1,1} - \phi_1^i [\beta_1^i x_{t-2,1} - \log(y_{t-1})] - \phi_2^i [\beta_1^i x_{t-3,1} - \log(y_{t-2})], \\ \text{logit}(\nu_t) &= \omega_0^i + \omega_1^i y_{t-1}, \quad i = 1, \dots, 1000.\end{aligned}$$

Ao realizar a previsão para o mês de outubro de 2022 obteve-se que não são esperados óbitos de idosos em decorrência de influenza, dado que o valor esperado de  $y_{out} \approx 0$ . Para este mesmo mês, a probabilidade estimada de que não ocorram óbitos foi de 0,719, variando entre 0,644 e 0,797 conforme o intervalo de credibilidade ao nível de 95%. No entanto, a esperança do componente positivo, ou seja  $E(y_{out} | y_{out} > 0)$ , se aproxima de 1. Isso indica que caso ocorram óbitos em outubro de 2022, a contagem esperada é igual a um.

## 6. Considerações finais

Neste trabalho foi proposta uma abordagem GARMA(p, q) para a modelagem de séries temporais de contagem com excesso de zeros assumindo a distribuição Poisson zero-ajustada cujos parâmetros variam no tempo, possibilitando avaliar e realizar previsões de contagens iguais a zero através do parâmetro  $\nu_t$ . Adotou-se uma perspectiva via inferência Bayesiana utilizando o algoritmo HMC para a amostragem da posteriori conjunta, que pode apresentar benefícios em termos da exploração do espaço de parâmetros.

De modo a avaliar o desempenho inferencial, foi realizado um estudo de simulação considerando variações das ordens autorregressivas e da proporção de zeros. Os resultados obtidos computacionalmente indicaram um bom desempenho inferencial, sendo, em grande parte das vezes, aprimorados conforme o tamanho amostral cresceu. Em relação à aplicação do modelo aos dados de influenza, pôde-se verificar a aplicabilidade e utilidade do modelo ao possibilitar a previsão e a mensuração da probabilidade de ocorrência de óbitos.

## Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## Referências

- [1] M. Alqawba, N. Diawara, and N. Chaganty, *Zero-inflated count time series models using gaussian copula*, *Sequential analysis* 38 (2019), pp. 342–357.
- [2] B. Andrade, M. Andrade, and R. Ehlers, *Bayesian gamma models for count data*, *Communications in statistics: case studies, data analysis and applications* 1 (2015), pp. 192–205.
- [3] A. Aragaw, A. Azene, and M. Workie, *Poisson logit hurdle model with associated factors of perinatal mortality in ethiopia*, *Journal of big data* 9 (2022), pp. 1–11.
- [4] W. Barreto-Souza, *Mixed poisson inar(1) processes*, *Statistical papers* 60 (2017), p. 2119–2139.

- [5] M. Benjamin, R. Rigby, and M. Stasinopoulos, *Generalized autoregressive moving average models*, Journal of the American statistical association 98 (2003), pp. 214–223.
- [6] G. Box and G. Jenkins, *Time series analysis, forecasting and control*, Holden-Day, 1976.
- [7] G. Box and D. Pierce, *Distribution of residual autocorrelations in autoregressive-integrated moving average time series models*, Journal of the American statistical association 65 (1970), pp. 1509–1526.
- [8] O. Briet, P. Amerasinghe, and P. Vounatsou, *Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers*, Plos one 8 (2013), pp. 1–9.
- [9] L. Broemeling, *Bayesian analysis of time series*, CRC Press, 2019.
- [10] M. Burda and J. Maheu, *Bayesian adaptive hamiltonian monte carlo with an application to high-dimensional bekk garch models*, Studies in nonlinear dynamics and econometrics 17 (2013), pp. 345–372.
- [11] F. Canova and B. Hansen, *Are seasonal patterns constant over time? a test for seasonal stability*, Journal of business & economic statistics 13 (1995), p. 237–252.
- [12] K. Conceição, A. Suzuki, and M. Andrade, *A bayesian approach for zero-modified skellam model with hamiltonian mcmc*, Statistical methods & applications 30 (2021), pp. 747–765.
- [13] D. Cox and A. Stuart, *Some quick sign tests for trend in location and dispersion*, Biometrika 42 (1955), pp. 80–95.
- [14] R. Davis, K. Fokianos, S. Holan, H. Joe, J. Livsey, R. Lund, V. Pipiras, and N. Ravishanker, *Count time series: A methodological review*, Journal of the American statistical association 0 (2021), pp. 1–15.
- [15] Departamento de Informática do Sistema Único de Saúde, *Morbidade hospitalar do Sistema Único de Saúde* (2022). Available at <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sih/cnv/nibr.def>.
- [16] S. Duane, A. Kennedy, B. Pendleton, and D. Roweth, *Hybrid monte carlo*, Physics letters B 195 (1987), pp. 216–222.
- [17] P. Dunn and G. Smyth, *Randomized quantile residuals*, Journal of computational and graphical statistics 5 (1996), pp. 236–244.
- [18] R. Ehlers, *A conway-maxwell-poisson gamma model for count data*, Arxiv (2019), pp. 1–11.
- [19] C. Feng, *A comparison of zero-inflated and hurdle models for modeling zero-inflated count data*, Journal of statistical distributions and applications 8 (2021), pp. 1–19.
- [20] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian data analysis* (2014).
- [21] M. Ghahramani and S. White, *Time series regression for zero-inflated and overdispersed count data: a functional response model approach*, Journal of statistical theory and practice 14 (2020), pp. 1–18.
- [22] L. Hashim, K. Hashim, and M. Shiker, *An application comparison of two poisson models on zero count data*, Journal of physics 1818 (2021).
- [23] M. Hoffman and A. Gelman, *The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo*, Journal of machine learning research 15 (2014), pp. 1593–1623.
- [24] I. Khandelwal, R. Adhikari, and G. Verma, *Time series forecasting using hybrid arima and ann models based on dwt decomposition*, Procedia computer science 48 (2015), pp. 173–179.
- [25] R. McElreath, *Statistical rethinking: A Bayesian course with examples in R and Stan*, Chapman & Hall, 2020.
- [26] M. Melo and A. Alencar, *Conway-maxwell-poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data*, Journal of time series analysis 41 (2020), pp. 830–857.
- [27] R. Neal, *Mcmc using hamiltonian dynamics*, in *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. Jones, and X. Meng, eds., chap. 5, Chapman & Hall, 2011, pp. 1–51.
- [28] E. Payne, J. Hardin, L. Egede, V. Ramakrishnan, A. Selassie, and M. Gebregziabher, *Approaches for dealing with various sources of overdispersion in modeling count data: scale adjustment versus modeling*, Statistical methods in medical research 26 (2017), pp.

- 1802–1823.
- [29] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2022). Available at <https://www.R-project.org/>.
  - [30] R. Rigby, M. Stasinopoulos, G. Heller, and F. Bastiani, *Distributions for modeling location, scale, and shape: using GAMLSS in R*, Chapman & Hall, 2019.
  - [31] A. Rocha and F. Cribari-Neto, *Beta autoregressive moving average models*, *Test* 18 (2008), pp. 529–545.
  - [32] T. Sáfadi and P. Morettin, *A bayesian analysis of autoregressive models with random normal coefficients*, *Journal of statistical computation and simulation* 73 (2003), pp. 563–573.
  - [33] L. Sales, A. Alencar, and L. Ho, *The berg generalized autoregressive moving average model for count time series*, *Computers & industrial engineering* 168 (2022), pp. 1–13.
  - [34] V. Sathish, S. Mukhopadhyay, and R. Tiwari, *Autoregressive and moving average models for zero-inflated count time series*, *Statistica Neerlandica* 76 (2021), pp. 1–23.
  - [35] R. Silva, *Generalized autoregressive neural network models*, ArXiv (2020).
  - [36] Stan Development Team, *RStan: the R interface to Stan*. Available at <https://mc-stan.org/>, R package version 2.26.13.
  - [37] M. Stasinopoulos and R. Rigby, *Gamlss.dist: distributions for generalized additive models for location scale and shape* (2020). Available at <https://CRAN.R-project.org/package=gamlss.dist>, R package version 5.1-6.
  - [38] M. Stasinopoulos, B. Rigby, and P. Eilers, *gamlss.util: GAMLSS utilities* (2016). Available at <https://CRAN.R-project.org/package=gamlss.util>, R package version 4.3-4.
  - [39] K. Tawiah, W. Iddrisu, and K. Asosega, *Zero-inflated time series modelling of covid-19 deaths in ghana*, *Journal of environmental and public health* (2021), pp. 1–9.
  - [40] G. Tian, Y. Liu, M. Tang, and X. Jiang, *Type i multivariate zero-truncated/adjusted poisson distributions with applications*, *Journal of computational and applied mathematics* 344 (2018), pp. 132–153.

## APÊNDICE A - Códigos

Estão apresentados neste apêndice os códigos em R utilizados neste trabalho. São códigos exemplificativos, portanto, os parâmetros, configurações e estruturas de dependência temporal podem ser alterados.

### Artigo 1

```
#=====
## Autores: Luiz Otávio Pala e Thelma Sáfadi
## 28 de abril de 2023
#=====
# ARTIGO 1
#-----
# Pacotes necessários
rm(list=ls(all=T))
library("gamlss.dist")
library("MHadaptive")
library("coda")
#-----
## Construindo o modelo com resposta Poisson

## Definindo os parâmetros para gerar  $yt \sim Poi(\mu)$ 
beta0 <- 1
phi1 <- 0.3
n <- 200

## Gerando  $yt$  de tamanho  $n + 5$  e removendo as 5 primeiras obs.
n <- n + 5
y <- c()
y[1] <- rpois(1, exp(beta0 + phi1*rnorm(1)))
for(i in 2:n){
  y[i] <- rpois(1, exp(beta0 + phi1*(log(max(y[i-1],0.1))-beta0)))
}
```

```

    }
y    <- y[6:n]

## Criando um dataframe para o processo simulado
data <- data.frame(y)
N    <- length(y)

## Função de verossimilhança aproximada
li_reg <- function(pars,data){
  beta0 <- pars[1]
  phil  <- pars[2]
  mu <- c()
  lv <- c()
  for(i in 2:N){
    mu[i] <- exp(beta0+phil*(log(max(data[i-1,1], 0.1))-beta0))
    lv[i] <- (dpois(data[i,1], mu[i], log = T))
  }
  log_likelihood <-sum(lv[2:N])
  prior          <- prior_reg(pars)
  return(log_likelihood + prior)
}

## Definindo as priors
prior_reg<-function(pars){
  beta0 <- pars[1]
  phil  <- pars[2]
  prior_beta0 <- dnorm(beta0, 0, 100, log = TRUE)
  prior_phil  <- dnorm(phil, 0, 100, log = TRUE)
  return(prior_phil + prior_beta0)
}

## MCMC

```



```

mcmc_r <- Metro_Hastings(li_func = li_reg, pars=c(0.9, 0.3),
                        par_names = c("beta0", "phil"),
                        data=data, burn_in = 1000,
                        iterations = 40000)

## Thin
mcmc_r <- mcmc_thin(mcmc_r, thin = 4)

## Análise de convergência
coda::raftery.diag(mcmc_r$trace)
coda::heidel.diag(mcmc_r$trace)
coda::geweke.diag(mcmc_r$trace)

## Estimativas a posteriori
(coef <- (apply(mcmc_r$trace, 2, mean)))
names(coef) <- c("beta0", "phil")
round(coef, 4)

#-----
## Construindo o modelo com resposta Binomial negativa

# Pacotes necessários
rm(list=ls(all=T))
library("gamlss.dist")
library("MHadaptive")
library("coda")

## Definindo os parâmetros para gerar  $yt \sim NBI(\mu, \sigma)$ 
beta0 <- 1
phil <- 0.3
sigma <- 0.25
n <- 200

```

```

## Gerando yt de tamanho n + 5 e removendo as 5 primeiras obs.
n    <- n + 5
y    <- c()
y[1] <- rNBI(1, exp(beta0 + phi1*rnorm(1)), sigma)
for(i in 2:n){
  y[i] <- rNBI(1, exp(beta0 + phi1*(log(max(y[i-1],0.1))-beta0)),sigma)
}
y    <- y[6:n]

## Criando um dataframe para o processo simulado
data <- data.frame(y)
N    <- length(y)

## Função de verossimilhança aproximada
li_reg <- function(pars,data){
  beta0 <- pars[1]
  phi1  <- pars[2]
  sigma <- pars[3]
  if(sigma<=0){return(NaN)}
  mu <- c()
  lv <- c()
  for(i in 2:N){
    mu[i] <- exp(beta0+phi1*(log(max(data[i-1,1], 0.1))-beta0))
    lv[i] <- (dNBI(data[i,1], mu[i], sigma, log = T))
  }
  log_likelihood <-sum(lv[2:N])
  prior          <- prior_reg(pars)
  return(log_likelihood + prior)
}

## Definindo as prioris
prior_reg<-function(pars)

```

```

{
  beta0 <- pars[1]
  phil  <- pars[2]
  sigma <- pars[3]
  prior_beta0 <- dnorm(beta0, 0, 100, log = TRUE)
  prior_phil  <- dnorm(phil, 0, 100, log = TRUE)
  prior_sigma <- dgamma(sigma, 1, 1/100, log = TRUE)
  return(prior_sigma + prior_phil + prior_beta0)
}

## MCMC
mcmc_r <- Metro_Hastings(li_func = li_reg,
                        pars=c(1.1, 0.1, 0.2),
                        par_names = c("beta0", "phil", "sigma"),
                        data=data, burn_in = 1000,
                        iterations = 40000)

## Thin
mcmc_r <- mcmc_thin(mcmc_r, thin = 4)

## Análise de convergência
coda::raftery.diag(mcmc_r$trace)
coda::heidel.diag(mcmc_r$trace)
coda::geweke.diag(mcmc_r$trace)

## Estimativas a posteriori
(coef <- (apply(mcmc_r$trace, 2, mean)))
names(coef) <- c("beta0", "phil", "sigma")
round(coef, 4)

#-----
## Construindo o modelo com resposta Poisson inversa Gaussiana

```

```
# Pacotes necessários
rm(list=ls(all=T))
library("gamlss.dist")
library("MHadaptive")
library("coda")

## Definindo os parâmetros para gerar  $y_t \sim \text{PIG}(\mu, \sigma)$ 
beta0 <- 1
phi1 <- 0.3
sigma <- 0.25
n <- 200

## Gerando  $y_t$  de tamanho  $n + 5$  e removendo as 5 primeiras obs.
n <- n + 5
y <- c()
y[1] <- rPIG(1, exp(beta0 + phi1*rnorm(1)), sigma)
for(i in 2:n){
  y[i] <- rPIG(1, exp(beta0 + phi1*(log(max(y[i-1],0.1))-beta0)), sigma)
}
y <- y[6:n]

# Criando um data frame para o processo simulado
data <- data.frame(y)
N <- length(y)

## Função de verossimilhança aproximada
li_reg <- function(pars,data){
  beta0 <- pars[1]
  phi1 <- pars[2]
  sigma <- pars[3]
  if(sigma<=0){return(NaN)}
}
```

```

mu <- c()
lv <- c()
for(i in 2:N){
  mu[i] <- exp(beta0+phil*(log(max(data[i-1,1], 0.1))-beta0))
  lv[i] <- (dPIG(data[i,1], mu[i], sigma, log = T))
}
log_likelihood <-sum(lv[2:N])
prior          <- prior_reg(pars)
return(log_likelihood + prior)
}

## Definindo as priors
prior_reg<-function(pars)
{
  beta0 <- pars[1]
  phil  <- pars[2]
  sigma <- pars[3]
  prior_beta0 <- dnorm(beta0, 0, 100, log = TRUE)
  prior_phil  <- dnorm(phil, 0, 100, log = TRUE)
  prior_sigma <- dgamma(sigma,1,1/100,log = TRUE)
  return(prior_sigma + prior_phil + prior_beta0)
}

## MCMC
mcmc_r <- Metro_Hastings(li_func = li_reg,
                        pars = c(1.5, 0.6, 0.01),
                        par_names = c("beta0", "phil", "sigma"),
                        data=data, burn_in = 1000,
                        iterations = 40000)

## Thin
mcmc_r <- mcmc_thin(mcmc_r, thin = 4)

```

```
## Análise de convergência
coda::raftery.diag(mcmc_r$trace)
coda::heidel.diag(mcmc_r$trace)
coda::geweke.diag(mcmc_r$trace)

## Estimativas a posteriori
(coef <- (apply(mcmc_r$trace, 2, mean)))
names(coef) <- c("beta0", "phi1", "sigma")
round(coef, 4)

#=====
# Fim.
#=====
```

## Artigo 2

```

#=====
## Autores: Luiz Otávio Pala e Thelma Sáfyadi
## 28 de abril de 2023
#=====
# ARTIGO 2
#-----
# Pacotes necessários
rm(list=ls(all=T))
library("gamlss.dist")
library("MHadaptive")
library("coda")
#-----
## Construindo o modelo com resposta Poisson zero-ajustada

## Definindo os parâmetros para gerar  $y_t \sim ZAP(\mu, \nu)$ 
beta0 <- 2
phi1 <- 0.2
nu <- 0.1
n <- 100

## Gerando  $y_t$  de tamanho  $n + 5$  e removendo as 5 primeiras obs.
n <- n + 5
y <- c()
y[1] <- rZAP(1, exp(beta0 + phi1*rnorm(1)), nu)
for(i in 2:n){
  y[i] <- rZAP(1, exp(beta0 + phi1*(log(max(y[i-1],0.1)) - beta0)), nu)
}
y <- y[6:n]

## Criando um dataframe para o processo simulado
data <- data.frame(y)

```

```

N    <- length(y)

## Função de verossimilhança aproximada
li_reg <- function(pars,data){
  beta0 <- pars[1]
  phil  <- pars[2]
  nu    <- pars[3]
  if(nu<= 0 | nu >= 1){return(NaN)}
  mu <- c()
  lv <- c()
  for(i in 2:N){
    mu[i] <- exp(beta0 + phil*(log(max(data[i-1, 1], 0.1)) - beta0))
    lv[i] <- (dZAP(data[i,1], mu[i], nu, log = T))
  }
  log_likelihood <-sum(lv[2:N])
  prior          <- prior_reg(pars)
  return(log_likelihood + prior)
}

## Definindo as priors
prior_reg<-function(pars)
{
  beta0 <- pars[1]
  phil  <- pars[2]
  nu    <- pars[3]

  prior_beta0 <- dnorm(beta0, 0, 100, log = TRUE)
  prior_phil  <- dnorm(phil, 0, 100, log = TRUE)
  prior_nu    <- dunif(nu,0, 1, log = T)
  return(prior_phil + prior_beta0 + prior_nu)
}

```



```

## MCMC
mcmc_r <- Metro_Hastings(li_func = li_reg, pars=c(1.2, 0.2, 0.2),
                        par_names = c("beta0", "phil", "nu"),
                        data = data, burn_in = 1000,
                        iterations = 30000)

## Thin
mcmc_r <- mcmc_thin(mcmc_r, thin = 4)

## Análise de convergência
coda::raftery.diag(mcmc_r$trace)
coda::heidel.diag(mcmc_r$trace)
coda::geweke.diag(mcmc_r$trace)

## Estimativas a posteriori
(coef <- (apply(mcmc_r$trace, 2, mean)))
names(coef) <- c("beta0", "phil", "nu")
round(coef, 4)

#-----
## Construindo o modelo com resposta Binomial neg. zero-ajustada

# Pacotes necessários
rm(list=ls(all=T))
library("gamlss.dist")
library("MHadaptive")
library("coda")

## Definindo os parametros para gerar  $yt \sim ZANBI(\mu, \nu, \sigma)$ 
beta0 <- 2
phil <- 0.3
nu <- 0.1
sigma <- 0.1

```

```

n      <- 100

## Gerando yt de tamanho n + 5 e removendo as 5 primeiras obs.
n      <- n + 5
y      <- c()
y[1] <- rZANBI(1, exp(beta0 + phi1*rnorm(1)), sigma, nu)
for(i in 2:n){
  y[i] <- rZANBI(1, exp(beta0 + phi1*(log(max(y[i-1],0.1))-beta0)),
                sigma = sigma, nu = nu)
}
y      <- y[6:n]

## Criando um dataframe para o processo simulado
data <- data.frame(y)
N      <- length(y)

## Função de verossimilhança aproximada
li_reg <- function(pars, data){
  beta0 <- pars[1]
  phi1  <- pars[2]
  sigma <- pars[3]
  nu    <- pars[4]
  if(nu<= 0 | nu >= 1){return(NaN)}
  if(sigma<= 0){return(NaN)}
  mu <- c()
  lv <- c()
  for(i in 2:N){
    mu[i] <- exp(beta0+phi1*(log(max(data[i-1,1], 0.1))-beta0))
    lv[i] <- (dZANBI(data[i,1], mu[i], sigma, nu, log = T))
  }
  log_likelihood <-sum(lv[2:N])
  prior          <- prior_reg(pars)

```

```

    return(log_likelihood + prior)
}

## Definindo as priors
prior_reg<-function(pars)
{
  beta0 <- pars[1]
  phi1  <- pars[2]
  sigma <- pars[3]
  nu    <- pars[4]
  prior_beta0 <- dnorm(beta0, 0, 100, log = TRUE)
  prior_phi1  <- dnorm(phi1, 0, 100, log = TRUE)
  prior_sigma <- dgamma(sigma,1,1/100,log = TRUE)
  prior_nu    <- dunif(nu,0,1, log = T)
  return(prior_phi1 + prior_beta0 + prior_sigma + prior_nu)
}

## MCMC
mcmc_r <- Metro_Hastings(li_func = li_reg,pars=c(1.2, 0.1, 0.1, 0.2),
                        par_names = c("beta0", "phi1", "sigma", "nu"),
                        data = data,
                        burn_in  = 1000, iterations = 40000)

## Thin
mcmc_r <- mcmc_thin(mcmc_r, thin = 4)

## Análise de convergência
coda::raftery.diag(mcmc_r$trace)
coda::heidel.diag(mcmc_r$trace)
coda::geweke.diag(mcmc_r$trace)

## Estimativas a posteriori
(coef <- (apply(mcmc_r$trace, 2, mean)))

```

```

names(coef) <- c("beta0", "phi1", "sigma", "nu")
round(coef, 4)

#-----
## Construindo o modelo com resp. Poisson inv. Gauss. zero-aj.

# Pacotes necessários
rm(list=ls(all=T))
library("gamlss.dist")
library("MHadaptive")
library("coda")

## Definindo os parâms. para gerar  $y_t \sim \text{ZAPIG}(\mu, \sigma, \nu)$ 
beta0 <- 2
phi1 <- 0.3
nu <- 0.1
sigma <- 0.1
n <- 100

## Gerando  $y_t$  de tamanho  $n + 5$  e removendo as 5 primeiras obs.
n <- n + 5
y <- c()
y[1] <- rZAPIG(1, exp(beta0 + phi1*rnorm(1)), sigma, nu)
for(i in 2:n){
  y[i] <- rZAPIG(1, exp(beta0 + phi1*(log(max(y[i-1],0.1)) - beta0)),
               sigma = sigma, nu = nu)
}
y <- y[6:n]

## Criando um dataframe para o processo simulado
data <- data.frame(y)
N <- length(y)

```

```

## Função de verossimilhança aproximada
li_reg <- function(pars,data){
  beta0 <- pars[1]
  phi1 <- pars[2]
  sigma <- pars[3]
  nu <- pars[4]
  if(nu<= 0 | nu >= 1){return(NaN)}
  if(sigma<= 0){return(NaN)}
  mu <- c()
  lv <- c()
  for(i in 2:N){
    mu[i] <- exp(beta0 + phi1*(log(max(data[i-1,1], 0.1)) - beta0))
    lv[i] <- (dZAPIG(data[i,1], mu[i], sigma, nu, log = T))
  }
  log_likelihood <-sum(lv[2:N])
  prior <- prior_reg(pars)
  return(log_likelihood + prior)
}

## Definindo as priors
prior_reg <- function(pars)
{
  beta0 <- pars[1]
  phi1 <- pars[2]
  sigma <- pars[3]
  nu <- pars[4]
  prior_beta0 <- dnorm(beta0, 0, 100, log = TRUE)
  prior_phi1 <- dnorm(phi1, 0, 100, log = TRUE)
  prior_sigma <- dgamma(sigma, 1, 1/100, log = TRUE)
  prior_nu <- dunif(nu,0,1, log = T)
  return(prior_phi1 + prior_beta0 + prior_sigma + prior_nu)
}

```

```
## MCMC
mcmc_r <- Metro_Hastings(li_func = li_reg, pars=c(1.2, 0.1, 0.1, 0.2),
                        par_names = c("beta0", "phil", "sigma", "nu"),
                        data = data, burn_in = 1000, iterations = 40000)

## Thin
mcmc_r <- mcmc_thin(mcmc_r, thin = 4)

## Análise de convergência
coda::raftery.diag(mcmc_r$trace)
coda::heidel.diag(mcmc_r$trace)
coda::geweke.diag(mcmc_r$trace)

## Estimativas a posteriori
(coef <- (apply(mcmc_r$trace, 2, mean)))
names(coef) <- c("beta0", "phil", "sigma", "nu")
round(coef, 4)

#=====
# Fim.
#=====
```

### Artigo 3

```

#=====
## Autores: Luiz Otávio Pala e Thelma Sáfadi
## 28 de abril de 2023
#=====
# ARTIGO 3
#-----
# Pacotes necessários
rm(list=ls(all=T))
library("rstan")
library("gamlss.dist")
library("arm")
rstan_options(auto_write = TRUE)
options(warn = -1)

## Definindo os parâmetros para gerar  $y_t \sim ZAP(\mu_t, \nu_t)$ 
w0 <- -2.0
w1 <- -1.0
phi1 <- 0.3
phi2 <- -0.2
n <- 200

## Gerando  $y_t$  de tamanho  $n + 5$  e removendo as 5 primeiras obs.
y <- c()
min <- 3
mu <- c()
nu <- c()
N <- n + min
y[1:min] <- 0
for(t in min:N){
  mu[t] <- exp(phi1*(log(max(y[t-1], 0.1))) + phi2*(log(max(y[t-2], 0.1))))
  nu[t] <- invlogit(w0 + w1*y[t-1])
}

```

```

    y[t] <- rZAP(1, mu[t], nu[t])
  }
y <- y[(min + 1):N]

## Criando uma lista para o processo simulado
data <- list(N = length(y), y = y)

## Definindo o modelo
model <- "
data {
  int<lower=0> N;
  int<lower=0> y[N];
}
parameters {
  real w0;
  real w1;
  real phi1;
  real phi2;
}

model {
  w0 ~ normal(0, 10);
  w1 ~ normal(0, 10);
  phi1 ~ normal(0, 10);
  phi2 ~ normal(0, 10);
for (n in 3:N) {
  if (y[n] == 0)
    target += log(inv_logit(w0 + w1*y[n-1]));
  else
    target += log1m(inv_logit(w0 + w1*y[n-1]))+
      poisson_lpmf(y[n] | exp(phi1*(log(fmax(y[n-1], 0.1))) +
        phi2*(log(fmax(y[n-2], 0.1))))) -

```



```

        poisson_lccdf(0 | exp(phi1*(log(fmax(y[n-1], 0.1))) +
        phi2*(log(fmax(y[n-2], 0.1))));
    }
}"

## HMC
fit <- (stan(model_code = model, data = data, iter = 20000, chains = 2,
            algorithm = "HMC", thin = 1, warmup = 4000))

## Estimativas a posteriori
S <- summary(fit, probs = c(0.025,0.975))
coef <- S$summary[1:4,1]
names(coef) <- c("w0", "w1", "phi1", "phi2")
round(coef, 4)

## Uma opção interessante para a análise
library(shinystan)
shinystan::launch_shinystan(fit)

#=====
# Fim.
#=====

```