

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

Luiz Otávio de O. Pala 
Federal University of Lavras

Marcela de M. Carvalho 
Federal University of Lavras

Thelma Sáfadi 
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, 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 (τ_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 $\boldsymbol{\Omega} = \{\mu_t | \mu_t > 0\}$ 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 $\boldsymbol{\Omega} = \{\mu_t, \sigma | \mu_t, \sigma > 0\}$ 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 (σ) 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 $\mu_t \subseteq \boldsymbol{\Omega}$, 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

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

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

where $\Omega = \{\mu_t, \sigma \mid \mu_t, \sigma > 0\}$ 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, $E(y_t \mid \mathcal{F}_{t-1}) = \mu_t$ and $V(y_t \mid \mathcal{F}_{t-1}) = \mu_t(1 + \mu_t\sigma)$. 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 $L(\boldsymbol{\theta} \mid Y)p(\boldsymbol{\theta})$, 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 μ_j and τ_j the hyperparameters associated to β_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 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 β_j (a) and σ (b). Analogously, the behavior of the densities of ϕ_k and θ_l is similar to that shown in Figure 1(a).

Considering the algebraic complexity of the joint posterior distributions, namely $\pi(\boldsymbol{\theta} \mid Y)$, 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.

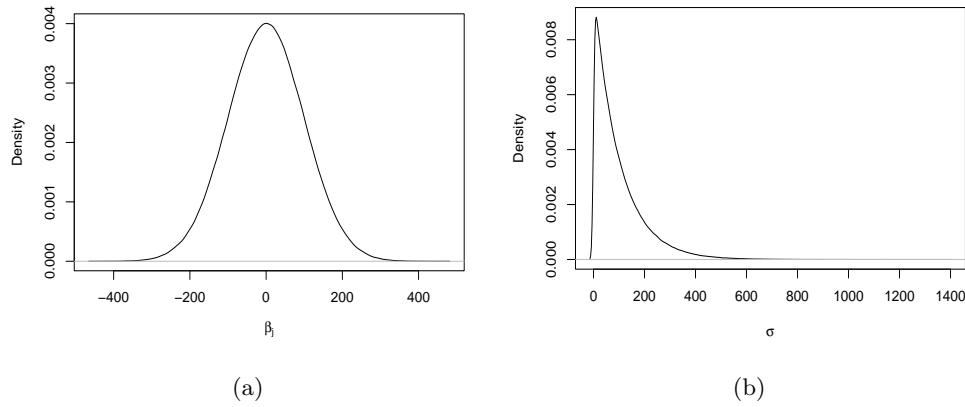


Figure 1: Prior densities of β_j and σ , respectively, that were used in the Bayesian analysis

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 μ_{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 α 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 $CB = \frac{1}{w} \sum_{i=1}^w \left| \frac{\hat{\theta} - \hat{\theta}^i}{\theta} \right|$, being τ the standard deviation of θ 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 $\boldsymbol{\beta} = (\beta_1)^\top$ is the vector associated with the level and $\boldsymbol{x}_t = (\boldsymbol{x}_1)$, being $\boldsymbol{x}_1 = (1, 1, \dots, 1)^\top$. 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 | β_1 | ϕ_1 | σ |
|----------|----------------|-----------|----------|----------|
| 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, σ , 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 σ , producing a good approximation when $n = 225$.

The other scenarios are included in the Appendix. When increasing ϕ_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 σ 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 | β_1 | 0.996 | 0.993 | 0.083 | 0.067 | 1.001 |
| | | ϕ_1 | 0.113 | 0.109 | 0.076 | 0.601 | 1.015 |
| | 125 | β_1 | 0.996 | 0.993 | 0.067 | 0.054 | 1.003 |
| | | ϕ_1 | 0.105 | 0.103 | 0.059 | 0.471 | 1.004 |
| | 225 | β_1 | 0.998 | 0.997 | 0.047 | 0.038 | 1.001 |
| | | ϕ_1 | 0.102 | 0.101 | 0.043 | 0.346 | 1.001 |
| NB-AR(1) | 75 | β_1 | 1.005 | 0.995 | 0.113 | 0.089 | 1.001 |
| | | ϕ_1 | 0.118 | 0.116 | 0.077 | 0.626 | 1.027 |
| | | σ | 0.309 | 0.271 | 0.121 | 0.403 | 1.113 |
| | 125 | β_1 | 0.997 | 0.992 | 0.084 | 0.067 | 1.000 |
| | | ϕ_1 | 0.109 | 0.108 | 0.059 | 0.467 | 1.011 |
| | | σ | 0.289 | 0.261 | 0.088 | 0.293 | 1.064 |
| | 225 | β_1 | 0.997 | 0.995 | 0.062 | 0.049 | 1.000 |
| | | ϕ_1 | 0.104 | 0.103 | 0.045 | 0.362 | 1.003 |
| | | σ | 0.266 | 0.255 | 0.066 | 0.216 | 1.029 |
| PIG-AR(1) | 75 | β_1 | 1.018 | 1.004 | 0.112 | 0.090 | 1.012 |
| | | ϕ_1 | 0.114 | 0.111 | 0.081 | 0.639 | 1.014 |
| | | σ | 0.348 | 0.278 | 0.158 | 0.541 | 1.176 |
| | 125 | β_1 | 1.005 | 0.998 | 0.087 | 0.070 | 1.001 |
| | | ϕ_1 | 0.107 | 0.107 | 0.060 | 0.473 | 1.006 |
| | | σ | 0.307 | 0.270 | 0.109 | 0.373 | 1.127 |
| | 225 | β_1 | 1.006 | 1.002 | 0.064 | 0.051 | 1.003 |
| | | ϕ_1 | 0.103 | 0.102 | 0.046 | 0.366 | 1.001 |
| | | σ | 0.275 | 0.256 | 0.074 | 0.245 | 1.053 |

Similar results were found in the second and third scenarios, where the estimates of ϕ_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 ϕ_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 $\text{thin} = 20$ 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:

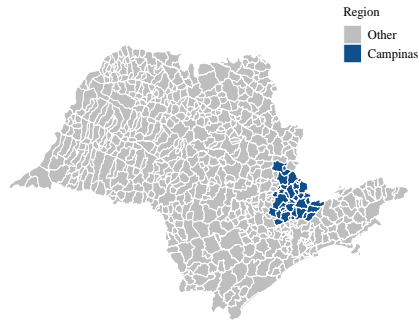


Figure 2: Geographic location of the cities included in the *Campinas* region. State of *São Paulo*, Brazil

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 ΔY (series adjusted for trend) is shown in Figure 3 (b) and 3 (c).

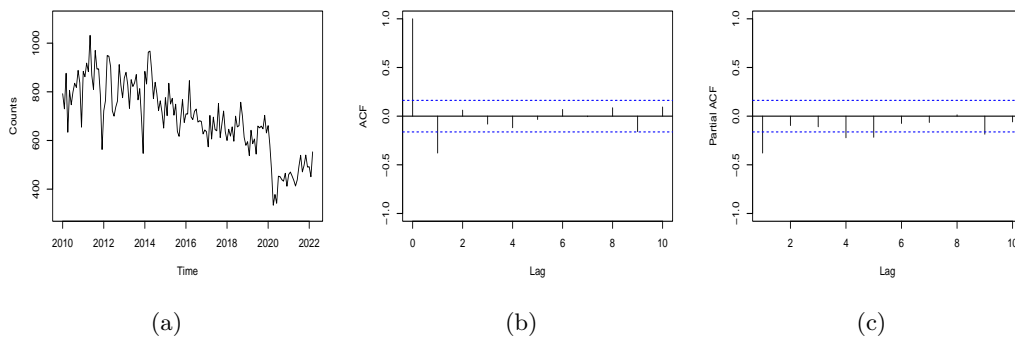


Figure 3: (a): Number of thefts reported in the region of *Campinas*; (b): Autocorrelation function of ΔY ; (c): Partial autocorrelation function of ΔY

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 β_2 of the POI-AR(1) and NB-AR(1) models. The greatest distances between the Bayesian and maximum likelihood estimates were observed in β_1 and ϕ_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) | β_1 | 6.8346 | 6.8342 | 0.0128 | 6.8107 | 6.8604 |
| | β_2 | -0.0042 | -0.0041 | 0.0002 | -0.0045 | -0.0038 |
| | ϕ_1 | 0.5246 | 0.5227 | 0.0241 | 0.4776 | 0.5716 |
| NB-AR(1) | β_1 | 6.8534 | 6.8546 | 0.0465 | 6.7604 | 6.9437 |
| | β_2 | -0.0044 | -0.0044 | 0.0006 | -0.0055 | -0.0033 |
| | ϕ_1 | 0.5663 | 0.5560 | 0.0706 | 0.4238 | 0.7010 |
| | σ | 0.0116 | 0.0114 | 0.0016 | 0.0087 | 0.0149 |
| PIG-AR(1) | β_1 | 6.8517 | 6.8503 | 0.0486 | 6.7525 | 6.9456 |
| | β_2 | -0.0044 | -0.0044 | 0.0006 | -0.0055 | -0.0033 |
| | ϕ_1 | 0.5816 | 0.5817 | 0.0723 | 0.4380 | 0.7192 |
| | σ | 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 β_1 and ϕ_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 | 4328.217 | 4332.140 |
| CPO | -1245.136 | -842.182 | -843.139 |

We proceeded with the analysis estimating the randomized quantile residuals proposed by Dunn and Smyth (1996), defined as $\mathbf{r}_t = \Phi^{-1}(F_{y_t}(y_t | \mathcal{F}_{t-1}))$, where Φ^{-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.

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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1080/03461238.1987.10413823).

A. Appendix

A.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 | β_1 | 0.978 | 0.974 | 0.137 | 0.108 | 1.012 |
| | | ϕ_1 | 0.376 | 0.370 | 0.080 | 0.166 | 1.044 |
| | 125 | β_1 | 0.987 | 0.984 | 0.101 | 0.078 | 1.008 |
| | | ϕ_1 | 0.395 | 0.392 | 0.065 | 0.131 | 1.002 |
| | 225 | β_1 | 0.993 | 0.992 | 0.076 | 0.061 | 1.004 |
| | | ϕ_1 | 0.400 | 0.399 | 0.051 | 0.102 | 1.000 |
| NB-AR(1) | 75 | β_1 | 0.990 | 0.959 | 0.185 | 0.149 | 1.001 |
| | | ϕ_1 | 0.389 | 0.385 | 0.084 | 0.171 | 1.007 |
| | | σ | 0.343 | 0.292 | 0.140 | 0.503 | 1.202 |
| | 125 | β_1 | 0.997 | 0.982 | 0.141 | 0.110 | 1.000 |
| | | ϕ_1 | 0.402 | 0.399 | 0.066 | 0.132 | 1.000 |
| | | σ | 0.293 | 0.265 | 0.103 | 0.339 | 1.083 |
| | 225 | β_1 | 0.996 | 0.988 | 0.102 | 0.082 | 1.000 |
| | | ϕ_1 | 0.401 | 0.400 | 0.050 | 0.101 | 1.000 |
| | | σ | 0.264 | 0.250 | 0.077 | 0.245 | 1.016 |
| PIG-AR(1) | 75 | β_1 | 1.042 | 1.000 | 0.172 | 0.141 | 1.029 |
| | | ϕ_1 | 0.390 | 0.384 | 0.086 | 0.174 | 1.007 |
| | | σ | 0.400 | 0.307 | 0.170 | 0.680 | 1.336 |
| | 125 | β_1 | 1.005 | 0.985 | 0.140 | 0.112 | 1.000 |
| | | ϕ_1 | 0.400 | 0.397 | 0.068 | 0.135 | 1.000 |
| | | σ | 0.319 | 0.271 | 0.128 | 0.435 | 1.136 |
| | 225 | β_1 | 1.006 | 0.997 | 0.105 | 0.083 | 1.001 |
| | | ϕ_1 | 0.401 | 0.400 | 0.049 | 0.098 | 1.000 |
| | | σ | 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 | β_1 | 0.970 | 0.965 | 0.272 | 0.208 | 1.006 |
| | | ϕ_1 | 0.610 | 0.602 | 0.067 | 0.133 | 1.682 |
| | 125 | β_1 | 0.967 | 0.968 | 0.218 | 0.171 | 1.011 |
| | | ϕ_1 | 0.671 | 0.666 | 0.057 | 0.070 | 1.124 |
| | 225 | β_1 | 0.963 | 0.965 | 0.173 | 0.140 | 1.022 |
| | | ϕ_1 | 0.697 | 0.695 | 0.050 | 0.058 | 1.001 |
| NB-AR(1) | 75 | β_1 | 1.088 | 0.979 | 0.381 | 0.312 | 1.026 |
| | | ϕ_1 | 0.627 | 0.617 | 0.074 | 0.118 | 1.402 |
| | | σ | 0.330 | 0.259 | 0.162 | 0.500 | 1.114 |
| | 125 | β_1 | 1.040 | 0.960 | 0.321 | 0.256 | 1.007 |
| | | ϕ_1 | 0.675 | 0.669 | 0.064 | 0.077 | 1.073 |
| | | σ | 0.308 | 0.265 | 0.123 | 0.404 | 1.105 |
| | 225 | β_1 | 0.986 | 0.940 | 0.264 | 0.207 | 1.001 |
| | | ϕ_1 | 0.697 | 0.694 | 0.055 | 0.063 | 1.001 |
| | | σ | 0.289 | 0.265 | 0.099 | 0.319 | 1.075 |
| PIG-AR(1) | 75 | β_1 | 1.163 | 1.016 | 0.366 | 0.322 | 1.094 |
| | | ϕ_1 | 0.618 | 0.607 | 0.072 | 0.125 | 1.508 |
| | | σ | 0.394 | 0.269 | 0.210 | 0.700 | 1.212 |
| | 125 | β_1 | 1.073 | 0.976 | 0.316 | 0.260 | 1.026 |
| | | ϕ_1 | 0.660 | 0.653 | 0.061 | 0.081 | 1.197 |
| | | σ | 0.363 | 0.286 | 0.167 | 0.576 | 1.207 |
| | 225 | β_1 | 1.027 | 0.972 | 0.256 | 0.208 | 1.005 |
| | | ϕ_1 | 0.700 | 0.696 | 0.053 | 0.061 | 1.000 |
| | | σ | 0.306 | 0.267 | 0.114 | 0.387 | 1.115 |

A.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} | θ | \overline{I} | \overline{G} | \overline{HW} |
|-----------|-------|---------|---------|------|-----------------|-----------|----------------|----------------|-----------------|
| POI-AR(1) | 75 | 5,000 | 1,000 | 7 | 0.54 | β_1 | 1.16 | 0.02 | 0.53 |
| | | | | | | ϕ_1 | 1.28 | 0.01 | 0.51 |
| | 125 | 5,000 | 1,000 | 7 | 0.54 | β_1 | 1.06 | 0.03 | 0.52 |
| | | | | | | ϕ_1 | 1.12 | 0.01 | 0.53 |
| | 225 | 5,000 | 1,000 | 7 | 0.51 | β_1 | 1.26 | 0.02 | 0.52 |
| | | | | | | ϕ_1 | 1.25 | 0.01 | 0.53 |
| NB-AR(1) | 75 | 5,000 | 1,000 | 7 | 0.43 | β_1 | 1.29 | 0.01 | 0.52 |
| | | | | | | ϕ_1 | 1.32 | 0.01 | 0.52 |
| | 125 | 5,000 | 1,000 | 8 | 0.44 | σ | 1.17 | 0.02 | 0.52 |
| | | | | | | β_1 | 1.24 | 0.01 | 0.53 |
| | 225 | 5,000 | 1,000 | 7 | 0.45 | ϕ_1 | 1.26 | 0.01 | 0.52 |
| | | | | | | σ | 1.14 | 0.06 | 0.51 |
| PIG-AR(1) | 75 | 5,000 | 1,000 | 8 | 0.42 | β_1 | 1.29 | 0.02 | 0.53 |
| | | | | | | ϕ_1 | 1.30 | 0.00 | 0.51 |
| | 125 | 5,000 | 1,000 | 8 | 0.43 | σ | 1.22 | 0.01 | 0.51 |
| | | | | | | β_1 | 1.24 | 0.01 | 0.52 |
| | 225 | 5,000 | 1,000 | 8 | 0.44 | ϕ_1 | 1.27 | 0.00 | 0.52 |
| | | | | | | σ | 1.10 | 0.07 | 0.53 |
| 225 | 5,000 | 1,000 | 8 | 0.44 | β_1 | 1.24 | 0.03 | 0.51 | |
| | | | | | ϕ_1 | 1.26 | 0.04 | 0.51 | |
| 225 | 5,000 | 1,000 | 8 | 0.44 | σ | 1.11 | 0.09 | 0.50 | |
| | | | | | β_1 | 1.23 | 0.01 | 0.50 | |
| 225 | 5,000 | 1,000 | 8 | 0.44 | ϕ_1 | 1.25 | 0.01 | 0.51 | |
| | | | | | σ | 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} | θ | \overline{I} | \overline{G} | \overline{HW} |
|-----------|-------|---------|---------|------|-----------------|-----------|----------------|----------------|-----------------|
| POI-AR(1) | 75 | 5,000 | 1,000 | 7 | 0.55 | β_1 | 1.29 | 0.01 | 0.52 |
| | | | | | | ϕ_1 | 1.23 | 0.01 | 0.52 |
| | 125 | 5,000 | 1,000 | 7 | 0.55 | β_1 | 1.28 | 0.01 | 0.51 |
| | | | | | | ϕ_1 | 1.23 | 0.05 | 0.51 |
| | 225 | 5,000 | 1,000 | 8 | 0.55 | β_1 | 1.20 | 0.03 | 0.52 |
| | | | | | | ϕ_1 | 1.18 | 0.03 | 0.51 |
| NB-AR(1) | 75 | 5,000 | 1,000 | 8 | 0.42 | β_1 | 1.25 | 0.01 | 0.52 |
| | | | | | | ϕ_1 | 1.26 | 0.00 | 0.53 |
| | | | | | | σ | 1.12 | 0.01 | 0.52 |
| | 125 | 5,000 | 1,000 | 8 | 0.44 | β_1 | 1.24 | 0.01 | 0.52 |
| | | | | | | ϕ_1 | 1.25 | 0.02 | 0.52 |
| | | | | | | σ | 1.13 | 0.06 | 0.51 |
| 225 | 5,000 | 1,000 | 8 | 0.44 | β_1 | 1.21 | 0.04 | 0.49 | |
| | | | | | ϕ_1 | 1.22 | 0.00 | 0.51 | |
| | | | | | σ | 1.13 | 0.12 | 0.50 | |
| PIG-AR(1) | 75 | 5,000 | 1,000 | 7 | 0.40 | β_1 | 1.29 | 0.07 | 0.51 |
| | | | | | | ϕ_1 | 1.32 | 0.05 | 0.53 |
| | | | | | | σ | 1.15 | 0.06 | 0.52 |
| | 125 | 5,000 | 1,000 | 8 | 0.43 | β_1 | 1.24 | 0.00 | 0.52 |
| | | | | | | ϕ_1 | 1.26 | 0.03 | 0.52 |
| | | | | | | σ | 1.11 | 0.08 | 0.51 |
| 225 | 5,000 | 1,000 | 7 | 0.44 | β_1 | 1.28 | 0.01 | 0.50 | |
| | | | | | ϕ_1 | 1.30 | 0.03 | 0.51 | |
| | | | | | σ | 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} | θ | \overline{I} | \overline{G} | \overline{HW} |
|-----------|--------|---------|---------|------|-----------------|-----------|----------------|----------------|-----------------|
| POI-AR(1) | 75 | 7,000 | 1,000 | 12 | 0.51 | β_1 | 1.29 | 0.04 | 0.51 |
| | | | | | | ϕ_1 | 1.30 | 0.01 | 0.52 |
| | 125 | 5,000 | 1,000 | 10 | 0.51 | β_1 | 1.24 | 0.03 | 0.51 |
| | | | | | | ϕ_1 | 1.30 | 0.00 | 0.50 |
| | 225 | 5,000 | 1,000 | 10 | 0.52 | β_1 | 1.25 | 0.01 | 0.52 |
| | | | | | | ϕ_1 | 1.20 | 0.02 | 0.51 |
| NB-AR(1) | 75 | 7,000 | 1,000 | 7 | 0.37 | β_1 | 1.30 | 0.15 | 0.52 |
| | | | | | | ϕ_1 | 1.34 | 0.04 | 0.51 |
| | | | | | | σ | 1.55 | 0.03 | 0.49 |
| | 125 | 7,000 | 1,000 | 8 | 0.39 | β_1 | 1.01 | 0.06 | 0.51 |
| | | | | | | ϕ_1 | 1.22 | 0.01 | 0.49 |
| | | | | | | σ | 1.23 | 0.01 | 0.49 |
| 225 | 10,000 | 1,000 | 8 | 0.41 | β_1 | 1.10 | 0.02 | 0.48 | |
| | | | | | ϕ_1 | 1.22 | 0.02 | 0.50 | |
| | | | | | σ | 1.23 | 0.01 | 0.49 | |
| PIG-AR(1) | 75 | 11,000 | 1,000 | 11 | 0.35 | β_1 | 1.32 | 0.17 | 0.53 |
| | | | | | | ϕ_1 | 1.25 | 0.08 | 0.50 |
| | | | | | | σ | 1.23 | 0.13 | 0.52 |
| | 125 | 5,000 | 1,000 | 12 | 0.39 | β_1 | 1.22 | 0.15 | 0.52 |
| | | | | | | ϕ_1 | 1.23 | 0.02 | 0.52 |
| | | | | | | σ | 1.14 | 0.13 | 0.51 |
| 225 | 5,000 | 1,000 | 12 | 0.40 | β_1 | 1.26 | 0.04 | 0.49 | |
| | | | | | ϕ_1 | 1.21 | 0.02 | 0.49 | |
| | | | | | σ | 1.29 | 0.01 | 0.49 | |

A.3. Behavior of the marginal densities

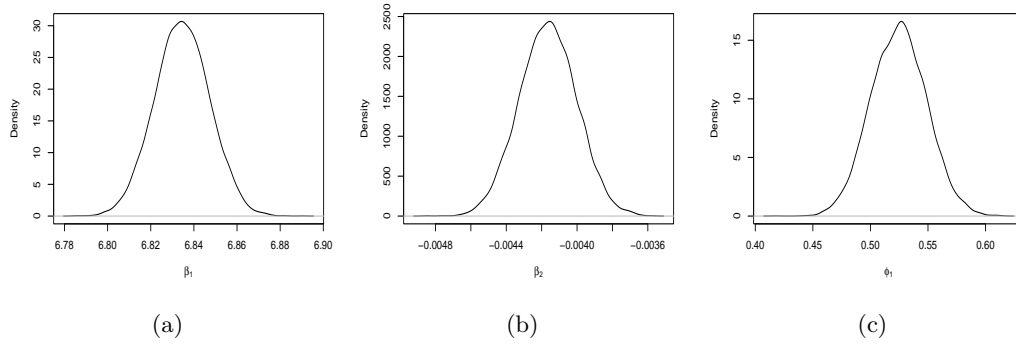


Figure 4: Posterior marginal densities of the POI-AR(1) parameters

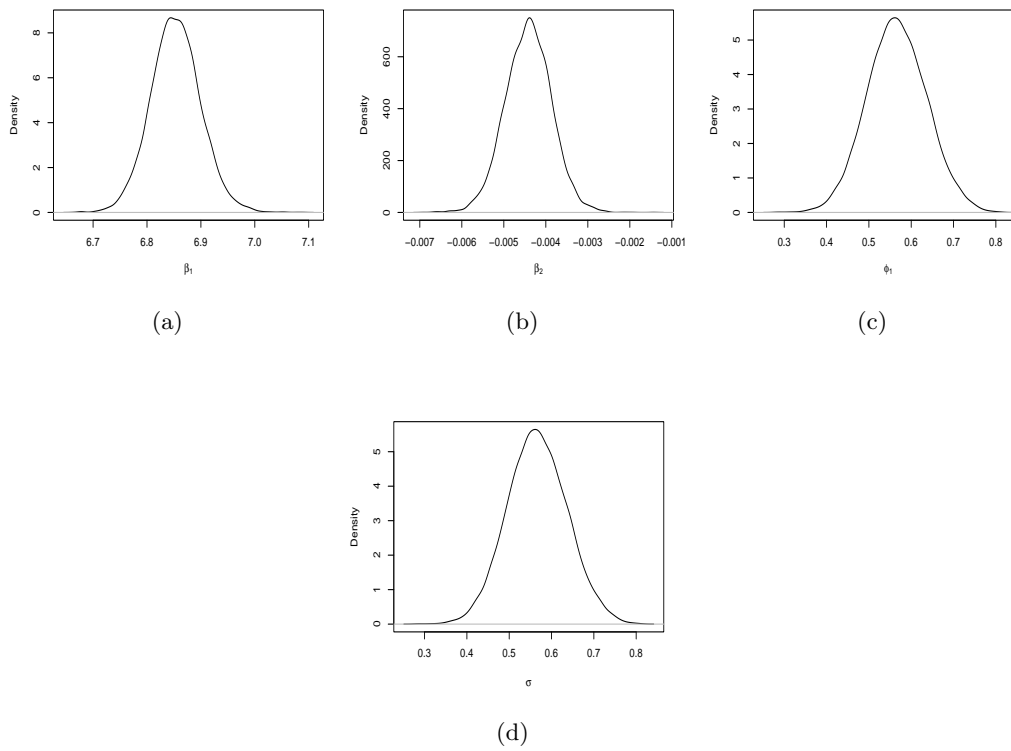


Figure 5: Posterior marginal densities of the NB-AR(1) parameters

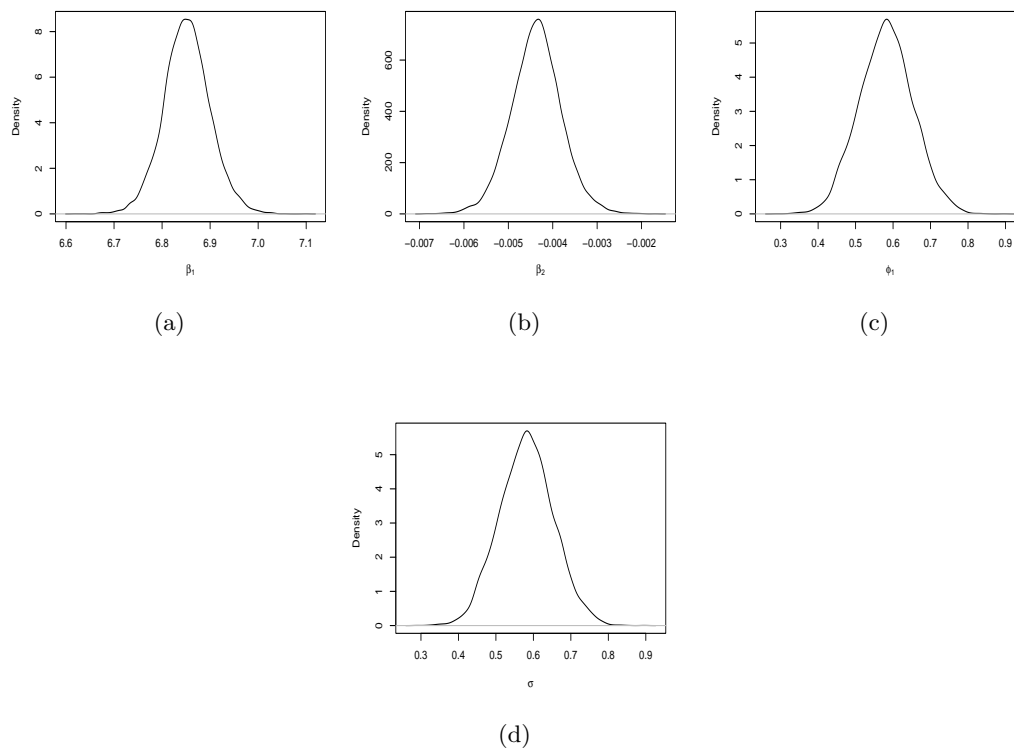


Figure 6: Posterior marginal densities of the FIG-AR(1) parameters

A.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) | β_1 | 6.8413 | 0.0128 |
| | β_2 | -0.0042 | 0.0002 |
| | ϕ_1 | 0.5235 | 0.0238 |
| NB-AR(1) | β_1 | 6.8578 | 0.0420 |
| | β_2 | -0.0044 | 0.0005 |
| | ϕ_1 | 0.5443 | 0.0673 |
| | σ | 0.0110 | 0.0014 |
| PIG-AR(1) | β_1 | 6.9841 | 0.1455 |
| | β_2 | -0.0058 | 0.0015 |
| | ϕ_1 | 0.6621 | 0.1460 |
| | σ | 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
 URL: <http://lattes.cnpq.br/5770791208849986>