A Monte Carlo study and data analysis for time series of counts

Glaura C. Franco,[a*] Helio S. Migon [b] Marcos Oliveira Prates [a]

[a] *Department of Statistics, UFMG, MG, Brazil.*
[b] *Department of Statistical Methods, UFRJ, RJ, Brazil*

Abstract

The objective of this paper is to present an empirical comparison between observation and parameter driven models using some procedures available in the literature that are commonly employed for time series of counts, based upon the conditional Poisson distribution. Although there are some works on this subject, as far as we know, an extensive simulation study comparing the two methodologies has not yet been performed. We aim to point out their main differences and similarities, concerning parameter estimation, model fitting and forecasting. The models are fitted under the Bayesian framework, using a Poisson generalized linear model with a latent AR(p) process in the mean, which accounts for autocorrelation in the data. In the parameter driven case, the INLA approach is used, while for the observation driven models addressed here, an MCMC procedure is developed to estimate the parameters. Concerning the covariates, coefficient estimates turn out to be remarkably similar across the different models. On the other hand, estimates for the autoregressive coefficients and the mean function and forecasts of future values depend heavily on the underlying process which generates the data. The results also show that when the parameters of the autoregressive process approach the non-stationary region, the parameter driven models present a much better performance.

Keywords: Poisson distribution, Observation driven model, Parameter driven model, autoregressive processes, Bayesian inference.

*Corresponding author: Glaura C. Franco. Address: Universidade Federal de Minas Gerais, Belo Horizonte, MG, 31270-901, Brazil. E-mail:glauraf@ufmg.br

1 Introduction

Models for non-Gaussian time series have been classified in observation driven and parameter driven models, for a general exponential family formulation (Cox, 1981). The main difference between the two approaches regards the way the dependence structure is incorporated to the model.

Let $\{y_t\}_{t=1,\ldots,n}$ be a time series and α_t a state variable, which can include unobserved components, covariates and noise variables. The two approaches can be illustrated by means of generalized state-space models, as pointed out in Davis et al. (1999). The observation equation in the state-space formulation is the same for the two approaches, and it gives the distribution of y_t conditional on a state variable, α_t.

The difference between observation and parameter driven models appears in the formulation of the state equation. In the observation driven case, the state equation is a function of past observations

$$\alpha_t = f(Y^{(t-1)}, \nu_t)$$

while in the parameter driven approach, it is a function of the latent process,

$$\alpha_t = h(\alpha^{(t-1)}, \varpi_t)$$

where superscripts denote past histories and ν_t and ϖ_t are white noise errors.

The basis for the observation driven models analyzed in this paper was first introduced in Zeger and Qaqish (1988). They proposed a quasi-likelihood approach to time series regression, focusing on Poisson and Gamma distributions. Later, Li (1994) extends the Zeger and Qaqish (1988) procedure by allowing a moving average structure to be added to the model. More recently, Benjamim et al. (2003) compiled the results of these two works in a more general formulation, which they call generalized autoregressive moving average (GARMA) models. The GARMA model can accommodate non stationary behavior and include a variety of different distributions, such as the Poisson, Binomial Logistic, Gamma and models of the GARCH class.

Following the same idea of Zeger and Qaqish (1988) but incorporating a mean correction component to generate stationary processes, Davis et al. (1999, 2003) propose an alternative observation driven model, called generalized linear autoregressive moving average (GLARMA) model. In this direction, they extend the work of Shephard (1995) by using standardized residuals with various powers of the conditional variance. The authors also provide maximum likelihood estimation and investigate the properties of their model, but they only consider the Poisson distribution.

Another group of observation driven models, recently proposed by Creal et al. (2013), is the generalized autoregressive score (GAS) model. The idea

of the GAS model is to use conditional observation densities to explain the dynamics of the time-varying parameter, which is updated on time via a scaled version of the score function.

Concerning parameter driven models, in the Bayesian context there is the seminal paper of West et al. (1985), who proposed a dynamic evolution structure to the parameters of an observation equation governed by distributions belonging to the exponential family. An important feature of their work is that the predictive distribution is obtained in closed form, by the use of appropriate conjugated priors. Following this line, some papers are devoted to the study of non-Gaussian models that possess conjugate filtering recursions, such as Smith and Miller (1986), Harvey and Fernandez (1989) and, more recently, Gamerman et al. (2013).

Considering the classical perspective, the first parameter driven model was proposed by Zeger (1988) for time series of counts. He incorporates latent processes in the conditional mean function of a log linear regression model which introduces both, overdispersion and autocorrelation, in the time series. Estimation of the latent process is based on a quasi-likelihood approach and requires methods based on Monte Carlo integration, as the likelihood cannot be written down in closed form. Some authors have applied MCMC methods to perform a Bayesian analysis for the Zeger model, such as Jung et al. (2006) and Czado and Kolbe (2004). State-space models with non-Gaussian observations were also considered in Shephard and Pitt (1997) and Durbin and Koopman (1997, 2000). The latter mix Kalman filter techniques and importance sampling to perform classical and Bayesian analysis.

As it can be seen, in the last two decades there has been a burst of new procedures in the area of non-Gaussian time series observations, either for parameter or observation driven models. But while there are many works devoted to this subject, little comparative studies have been performed so far. Some authors have attempted to compare the two classes, focusing on time series of counts, which have a wide field of practical application. Davis et al. (1999) made a review of the existing models at that time, aiming to provide statistical properties, model building and diagnosis for both classes, as well as to compare their performance on real data set. Jung et al. (2006) and Jung and Tremayne (2011) compare observation and parameter driven models based on real data applications. The conclusions they draw is that the two approaches fit equally well to the data set under study, giving similar estimates to the coefficient of the covariates. Feigin et al. (2008) compare the two approaches by measuring the over-dispersion and first-order autocorrelation of 13 count series data. They conclude that the parameter driven models present much more series with empirical moments falling within the theoretical regions built.

The purpose of the present paper is to give some contribution towards

understanding the main differences between the two approaches, the observation and parameter driven models, by means of a large simulation study. We implement the comparisons using count data with Poisson distribution, as this is a very useful model in practical situations. Although being a simple model, it evidences the main aspects of the two approaches, thus the comparisons are easily performed. In the parameter driven case, we adopt the procedure introduced in Zeger (1988) with an autoregressive latent process in the mean, while the GARMA and GLARMA models are employed in the observation driven class. The models addressed here were chosen due to the great similarity presented by the component which holds the autocorrelation structure of the process, thus making it easier to evaluate their characteristics.

The paper also provides contribution concerning inference under the Bayesian framework. In the parameter driven case, as observed by Jung et al. (2006), efficient estimation of the Zeger (1988) model is difficult to perform, as the likelihood function depends on high-dimensional integrals. We propose here to use the integrated nested Laplace approximation (INLA) approach (Rue et al. (2009)), which provides fast Bayesian inference for the marginal posterior densities of the hyperparameters and the latent components. Concerning observation driven models, the existing literature on GARMA and GLARMA models only deals with maximum likelihood estimation. Thus, we propose in this article a Markov Chain Monte Carlo (MCMC) procedure to evaluate the full posterior densities of the model parameters and variables.

All these topics are presented in the paper as follows: Section 2 outlines the parameter and observation driven models addressed here; Section 3 presents some considerations on the computational aspects of the INLA and MCMC algorithms; Section 4 discusses the results of the Monte Carlo experiments. Some conclusions are drawn in Section 5.

2 Models

This section presents the parameter and observation driven models discussed in this paper. The two approaches are illustrated using a generalized linear model with Poisson distribution. In an observation driven model, the dependence structure can be added to the mean function through past information of the observed series, as well as past and present information obtained from the covariates. In contrast, for a parameter driven model the dependency is introduced adding a latent dynamic process to the mean.

Let $\boldsymbol{x}$ be a $m \times 1$ vector of covariates and $\mathscr{H}_t = (Y^{(t-1)}, \boldsymbol{x}_1, ..., \boldsymbol{x}_t)$ be a σ-field generated by past values of the observed series, $Y^{(t-1)} = (y_1, ..., y_{t-1})$, and past and present values of the covariates. Assume that the conditional

distribution of each observation, y_t, $t = 1, ..., n$, given the previous information set, $\mathscr{H}_t$, is Poisson with mean μ_t. In the Poisson regression model, the canonical link function is the log-link, i.e., $\eta_t = \ln(\mu_t)$. Thus, the observation and state equations for this model can be written, respectively, as

$$y_t \mid \mathscr{H}_t \sim Poisson(\mu_t) \tag{1}$$

$$\eta_t = \ln(\mu_t) = \boldsymbol{x}_t^{'}\boldsymbol{\beta} + Z_t, \quad t = 1, ..., n, \tag{2}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_m)$ and the structure of Z_t depends on the adopted approach.

Apart from the Z_t component, the main differences between the two approaches are related to the estimation process and the ability to generate forecasts. According to Diebold and Schuermann (1996), observation driven models are easier to estimate because they are defined in terms of conditional densities that depend only on past observations, even though exact maximum likelihood estimation is still difficult to obtain. By the other hand, inference in parameter driven models requires a considerable computational effort. Regarding forecast of future values, observation driven models present a more natural way of obtaining the forecasts, since in the parameter driven models the latent process is not directly observable. The advantages of the parameter driven approach relies mainly on two aspects: the natural way to interpret the effect of covariates and the proof of asymptotic properties, which are easily demonstrated depending on the latent process assumed. Asymptotic properties of the observation driven models are very difficult to establish, especially those concerning ergodicity and stationarity (see Davis et al. (1999)).

In the next subsections we present the procedures in more detail.

2.1 Parameter driven model

The basis for the parameter driven model described here was first proposed by Zeger (1988). It states that, conditional on a latent stationary process, ϵ_t, and on $\boldsymbol{x}_t$, the distribution of y_t is assumed to be Poisson with mean $\mu_t = exp\{\boldsymbol{x}_t^{'}\boldsymbol{\beta}\}\epsilon_t$, i.e.,

$$y_t \mid \boldsymbol{x}_t, \epsilon_t \sim Poisson\big(exp\{\boldsymbol{x}_t^{'}\boldsymbol{\beta}\}\epsilon_t\big), \quad t = 1, ..., n.$$

Zeger (1988) imposes the restriction that $E(\epsilon_t) = 1$, so that the unconditional mean,

$$\mu_t = E(y_t) = E(E(y_t \mid \epsilon_t)) = exp(\boldsymbol{x}_t^{'}\boldsymbol{\beta})$$

does not depend on moments of the ϵ_t series. According to Zeger (1988), the latent process ϵ_t introduces both overdispersion and autocorrelation into y_t.

For the Poisson model, it is convenient to model the logarithm of ϵ_t, $ln(\epsilon_t)$, in order to ensure that the conditional mean of y_t is non-negative. With this reparameterization, $ln(\epsilon_t)$ should be taken as a stationary Gaussian process, so that a relationship between the autocovariance of $\ln(\epsilon_t)$ and ϵ_t exists (see Davis et al. (1999)).

Based on the above considerations, one possibility to model the latent process is to to assume that $Z_t = \ln(\epsilon_t)$ in (2) follows a stationary Gaussian autoregressive process of order p. Thus, the parameter driven model stated here, called from now on Zeger-ARp, is given by

$$y_t \mid \boldsymbol{x}_t, u_t \sim Poisson(\mu_t), \quad t = 1, ..., n, \tag{3}$$

$$\eta_t = \ln(\mu_t) = \boldsymbol{x}_t^{'}\boldsymbol{\beta} + Z_t \tag{4}$$

$$\Phi(B)Z_t = u_t, \quad u_t \sim i.i.d.N(0, \sigma_u^2), \tag{5}$$

where $\Phi(B) = (1 - \sum_{j=1}^{p} \phi_j B^j)$, B is the backshift operator, p is a positive integer and $\Phi(z)$, with a scalar z, is a polynomial with all roots outside the unit circle. The AR(1) structure has already been considered in the literature, for example in the works of Chan and Ledolter (1995), Jung et al. (2006) and Czado and Kolbe (2004).

2.2 Observation driven models

As described earlier in the paper, the literature of time series of counts presents several options to work in the class of observation driven models. Here we discuss two of these procedures, which have a similar structure compared to the parameter driven model described in Section 2.1. The procedures were proposed simultaneously by Benjamim et al. (2003) and by Davis et al. (2003). The similarities and differences between the two approaches are described in the next subsections in more detail.

2.2.1 The generalized linear autoregressive moving average (GLARMA) model

The GLARMA model was first proposed by Shephard (1995) and latter extended by Davis et al. (2003, 1999), to allow for a more general structure for

the error term and also to permit pure MA or pure AR, besides the mixed ARMA process.

Davis et al. (2003) provide the basis for the GLARMA model whose conditional distribution, given the previous information set, follows a Poisson distribution. In this case, the state equation is composed of a linear model for the explanatory variables plus an error term with a moving average structure,

$$\eta_t = \ln(\mu_t) = \boldsymbol{x}_t^{'}\boldsymbol{\beta} + \sum_{i=1}^{\infty} \tau_i e_{t-i}$$

where

$$e_t = (y_t - \mu_t)/\mu_t^{\lambda}, \quad \lambda \geq 0.$$

In analogy with the ARMA models of Box and Jenkins (1976), it is possible to write the infinite moving average term as an autoregressive-moving average filter, by letting

$$\sum_{i=1}^{\infty} \tau_i z^i = (1 - \sum_{i=1}^{p} \phi_i z^i)^{-1}(1 - \sum_{i=1}^{q} \theta_i z^i) - 1.$$

Thus, the Poisson-GLARMA(p,q) model can be written as,

$$y_t \mid \mathscr{H}_t \sim Poisson(\mu_t), \quad t = 1, ..., n, \tag{6}$$

$$\eta_t = \ln(\mu_t) = \boldsymbol{x}_t^{'}\boldsymbol{\beta} + Z_t \tag{7}$$

$$Z_t = \sum_{i=1}^{\infty} \tau_i e_{t-i} = \sum_{i=1}^{p} \phi_i (Z_{t-i} + e_{t-i}) + \sum_{i=1}^{q} \theta_i e_{t-i}, \tag{8}$$

$$e_t = (y_t - e^{\eta_t})e^{-\lambda \eta_t}. \tag{9}$$

Davis et al. (2003) establishes some properties for special cases of the GLARMA model. For example, the process η_t is ergodic when $\lambda = 1, q = 1$ and $\boldsymbol{x}_t^{'}\boldsymbol{\beta} = \boldsymbol{\beta}$. For other values of λ, they could not prove the uniqueness, or even the existence (if $\lambda < 1/2$), of a stationary distribution. They also show, for this simple case, that, for $\lambda = 1$, the range of η_t does not depend on the value of η_{t-1}, which is not true for other values of λ. Therefore, it can be perceived that the properties of the GLARMA model are severely dependent on the value of λ.

2.2.2 The generalized autoregressive moving average (GARMA) model

Benjamim et al. (2003) have compiled the results of Zeger and Qaqish (1988) and Li (1994) to propose a more general class of observation driven models, called GARMA models. Although their procedure can accommodate any model in which the conditional distribution of y_t given the previous information set $\mathscr{F}_t = (\mathscr{H}_t, \mu_{t-1}, ..., \mu_1)$ belongs to the exponential family, in this paper we present only the results for the Poisson-GARMA model, to keep in line with the other procedures addressed here.

Let y_t given $\mathscr{F}_t$ be Poisson distributed with mean μ_t,

$$f(y_t \mid \mathscr{F}_t) = exp\{y_t \ln(\mu_t) - \mu_t - \ln(y_t!)\}, \quad t = 1, ..., n.$$

Assuming the log-link function, Benjamim et al. (2003) propose a sum of two components for Z_t in (2). The first component represents the autoregressive term and consists of a function of the past values of y_t and $\boldsymbol{x}_t$. The second component represents the moving average term and can be calculated using any sort of residuals computed from the fitted model. In their work, they opt for residuals on the predicted scale, $\ln(y_t) - \eta_t$.

Thus, the Poisson-GARMA model is given by the following equations

$$y_t \mid \mathscr{F}_t \sim Poisson(\mu_t), \quad t = 1, ..., n, \tag{10}$$

$$\eta_t = \ln(\mu_t) = \boldsymbol{x}_t'\boldsymbol{\beta} + Z_t \tag{11}$$

$$Z_t = \sum_{j=1}^{p} \phi_j \big(\ln(y^*_{t-j}) - \boldsymbol{x}_{t-j}\boldsymbol{\beta} \big) + \sum_{j=1}^{q} \theta_j \big(\ln(y^*_{t-j}/\mu_{t-j}) \big), \tag{12}$$

where y^*_{t-j}=max(y_{t-j}, c) and $0 < c < 1$. This restriction is necessary because of the logarithm link function, as y_t can present zero values, which in this case are replaced by the threshold parameter c.

Benjamim et al. (2003) prove some properties of the GARMA model, such as stationary conditions of the marginal means and variances, but only for the identity link. For other link functions, they could only perform some Monte Carlo simulation to investigate the stationary regions, as it seems it is not possible to obtain the first two moments of the marginal distributions.

3 Estimation and computational aspects

Most of the works available in the literature use the maximum likelihood approach to estimate the parameters of the Zeger-ARp model and the GARMA

and GLARMA procedures. Some authors have tried a Bayesian approach for the Zeger-AR(1) model, such as the works of Jung et al. (2006) and Czado and Kolbe (2004) but, as far as the authors know, inference for the observation driven models addressed here was only performed under the classical perspective.

In this work, inference for the models described in Section 2 are performed under the Bayesian framework. The integrated nested Laplace approximation (INLA) approach and the Metropolis-Hastings algorithm are used, respectively, for estimation in the parameter and observation driven models.

3.1 Inference in the parameter driven model

Several methods to estimate parameters and latent variables in parameter driven models have been proposed in the literature. Zeger (1988) proposes estimating β and ϕ separately. Given consistent estimates for ϕ and σ^2, which can be achieved by a method of moments, the β's are estimated using quasi-likelihood. In another attempt to give efficient estimation for the parameters, Chan and Ledolter (1995) propose to use a Monte Carlo EM algorithm to obtain the maximum likelihood estimators (MLE). The procedure is computer intensive and not easy to implement, requiring the use of Gibbs sampler. Davis et al. (1999) use Taylor expansions to build an approximate likelihood, which is maximized through numerical algorithms. Another proposal is given in Jung et al. (2006). The authors employ efficient importance sampling to approximate the complete likelihood, under the classical and Bayesian framework.

Let $\psi = (\beta', \phi', \sigma_u^2)'$ and $Z = (Z_1, ..., Z_n)$. The likelihood function is given by

$$L(\psi; Y^{(n)}, Z) = \prod_{i=1}^{n} f(y_i \mid Z_i) L(\psi; Z)$$

$$\approx \mid V^{-1} \mid^{1/2} exp\Big\{ \sum_{i=1}^{n} (Z_i + \boldsymbol{x}_t^{'}\boldsymbol{\beta}) y_i - \sum_{i=1}^{n} exp(Z_i + \boldsymbol{x}_t^{'}\boldsymbol{\beta}) - \frac{1}{2} Z' V^{-1} Z \Big\}$$

where V is the covariance matrix of Z.

Since the likelihood function cannot be written down analytically, the estimation process should be done numerically. In this work we propose to use the INLA approach (Rue et al., 2009) to obtain efficient estimates for the parameter vector ψ.

The INLA approach relies on a hierarchical structure, where the response y_t is assumed independent, conditional on some latent field $\mathbf{w}$ and a vector of hyperparameters $\varsigma = (\phi', \sigma_u^2)$. These models are called latent Gaussian

models (LGM). The LGM class assumes that the observation (or response) variable have a likelihood whose mean, μ_t, is linked to a structured additive predictor, η_t. In our setup, the link function is $\ln(\mu_t) = \eta_t$, as can be seen in (4). The INLA method approximates the posteriors of interest with a close form expression, which provides a great gain in computational time if compared with MCMC methods. Moreover, INLA does not have problems of covergence and mixing, inherent to MCMC runs.

Clearly, the parameter driven models of Section 2.1 satisfy the hierarchical condition imposed by INLA. Another restricition is that INLA requires that the latent field is Gaussian, which is also satified by the assumption in (5). Therefore, these characteristics are clearly suitable for the use of the INLA framework. In our specific framework, the latent field is composed by the linear predictor and the regression coefficients, $w_t = (\eta_t, \beta)$, and the AR coefficients and the noise variance define the hyperparameter vector $\varsigma = (\phi', \sigma_u^2)$.

For the model proposed in (4) and (5) we are interested in the following posterior marginal

$$\pi(w_t|Y^{(n)}) = \int \pi(w_t|Y^{(n)}, \varsigma)\pi(\varsigma|Y^{(n)})d\varsigma, \tag{13}$$

$$\pi(\varsigma_k|Y^{(n)}) = \int \pi(\varsigma_k|Y^{(n)})d\varsigma_{-k}, \quad k = 1, ..., p+1 \tag{14}$$

where ς_{-k} denotes vector ς without the k^{th} component.

The posterior marginals presented in (13) and (14) are calculated based on a Laplace approximation to the full conditionals $\pi(\varsigma|Y^{(n)})$ and $\pi(w_t|Y^{(n)}, \varsigma)$, $t = 1, \ldots, n$, and numerical integration routines to integrate out the hyperparameters ς.

In order to do so, INLA first approximates $\pi(\varsigma|Y^{(n)})$, using a Gaussian approximation to the full conditional distribution of $\mathbf{w} = \{w_1, ..., w_t\}$, by a multivariate Gaussian density $\tilde{\pi}_G(\mathbf{w}|Y^{(n)}, \varsigma)$ evaluated at its mode $\mathbf{w}^\star(\varsigma)$. Then the posterior density of ς is approximated by using the Laplace approximation (Tierney and Kadane, 1986)

$$\tilde{\pi}(\varsigma|Y^{(n)}) \propto \left.\frac{\pi(\mathbf{w}, Y^{(n)}, \varsigma)}{\tilde{\pi}_G(\mathbf{w}|Y^{(n)}, \varsigma)}\right|_{\mathbf{w}=\mathbf{w}^\star(\varsigma)}. \tag{15}$$

The second step is to compute the Laplace approximation of $\pi(w_t|Y^{(n)}, \varsigma)$ for selected values of ς, which will be used to perform a numerical integration to obtain the posterior marginals of w_t presented in (13). The density

$\pi(w_t|Y^{(n)},\varsigma)$ is approximated by

$$\tilde{\pi}_{LA}(w_t|Y^{(n)},\varsigma) \propto \left.\frac{\pi(\mathbf{w},Y^{(n)},\varsigma)}{\tilde{\pi}_G(\mathbf{w}_{-t}|w_t,Y^{(n)},\varsigma)}\right|_{\mathbf{w}_{-t}=\mathbf{w}^{\star}_{-t}(w_t,\varsigma)}, \tag{16}$$

where $\mathbf{w}_{-t}$ denotes the vector $\mathbf{w}$ without the t^{th} component, $\tilde{\pi}_G(\mathbf{w}_{-t}|w_{ij},\mathbf{y},\varsigma)$ is the Gaussian approximation of $\pi(\mathbf{w}_{-t}|w_t,\mathbf{y},\varsigma)$, treating w_t as observed and $\mathbf{w}^{\star}_{-t}(w_t,\varsigma)$ is the mode of $\pi(\mathbf{w}_{-t}|x_t,Y^{(n)},\varsigma)$.

The approximation $\tilde{\pi}_{LA}(w_t|Y^{(n)},\varsigma)$ in (16) can be quite expensive, since it is necessary to recompute $\tilde{\pi}_G(\mathbf{w}_{-t}|w_t,Y^{(n)},\varsigma)$ for all w_t and ς. Rue et al. (2009) proposes two alternatives to obtain these full conditionals in a cheaper way. We focus our analysis in the simplified Laplace approximation defined as the thrid order Taylor expansion of $\tilde{\pi}_{LA}(w_t|Y^{(n)},\varsigma)$ around $w_t = \mu_t(\varsigma)$ and approximated by a skew-normal distribution (for more details see Rue et al., 2009).

Finally, the full posterior approximations obtained previously are combined and the marginal posterior densities of w_t and ς_k are obtained by numerically integrating out the irrelevant terms. Therefore, the marginal approximation of the latent variables can be obtained by

$$\pi(w_t|Y^{(n)}) = \int \pi(w_t|\mathbf{y},\varsigma)\pi(\varsigma|Y^{(n)})d\varsigma \approx \sum_l \tilde{\pi}(w_t|Y^{(n)},\varsigma_l)\tilde{\pi}(\varsigma_l|Y^{(n)},)\Delta_l,$$

which is evaluated using a finite sum on a set ς_l of grid points, with area weights l for $l = 1, 2, \ldots, L$. Rue et al. (2009) argue that because the points ς_l are selected in a regular grid, it is feasible to take all the area weights l to be equal. In a similar way, the posterior marginal of $\pi(\varsigma_l|Y^{(n)})$ is obtained.

Thus, INLA directly calculates the posterior marginal of interest making posterior inference over ψ very efficient.

3.2 Inference in the observation driven model

In the observation driven approach the only quantity to be estimated is the parameter vector $\delta = (\beta', \phi', \theta')'$. In the GARMA model, the log-likelihood function is conditional on the first $s > max(p,q)$ observations and on $\ln(y_t^*)$ for $t = 1, ..., max(p,q)$. For the GLARMA model, the likelihood function is computed by conditioning on pre-period values of $e_0, e_{-1}, ..., e_{1-q}$ being equal to zero.

For both approaches, in the case of a Poisson distribution, the log-likelihood function is given by

$$l(\delta; Y^{(n)}) \propto \sum_{t=s+1}^{n} (y_t \eta_t(\delta) - e^{\eta_t(\delta)}), \tag{17}$$

where η_t and Z_t are given, respectively, by (7) and (8) for the GLARMA model and by (11) and (12) for the GARMA model.

Benjamim et al. (2003) provide maximum likelihood estimation in the GARMA model using iteratively reweighted least squares (IRLS) and partial likelihood (see a detailed description of this concept in Kedem and Fokianos (2002)) for the case when $\boldsymbol{x}_t$ is stochastic. The maximization of $l(\delta; Y^{(n)})$ is done using the Fisher scoring algorithm. In the GLARMA model, Davis et al. (2005) provide estimation by maximum likelihood, using the Newton-Raphson algorithm. They compute the first and second derivatives of an approximate likelihood, which is calculated conditional on $e_s = 0$, for $s \leq 0$. They claim that it is also possible to estimate parameter λ using this approximated likelihood, but they warn that care is needed as the likelihood is quite flat in the direction of λ. In Davis et al. (2005), the authors show the asymptotic normality of the maximum likelihood estimator, for the case $\lambda = 1$.

In this work, the estimation is performed using MCMC algorithms. The Bayesian approach initially requires the specification of a prior distribution for the parameter vector δ. The prior distribution, denoted by $\pi(\delta)$, should reflect the previous knowledge about the parameters, and its determination is, very often, not an easy task. Once the data is observed, the prior distribution can be updated to obtain the posterior distribution of the parameter vector as

$$\pi(\delta \mid Y^{(n)}) \propto \exp(l(\delta); Y^{(n)})\pi(\delta),$$

where $l(\delta; Y^{(n)})$ is given by (17).

One possibility for $\pi(\delta)$ is to use normal priors with zero mean and variance σ^2 for ϕ_i, θ_j and β_k, for $i = 1, ..., p$, $j = 1, ..., q$ and $k = 1, ..., m$. Assuming independence among the parameters, the joint prior is

$$\pi(\delta) \propto \frac{1}{(\sigma_\phi)^p (\sigma_\theta)^q (\sigma_\beta)^m} \exp\Big[- \frac{\sum_{i=1}^{p} \phi_i^2}{2\sigma_\phi^2} - \frac{\sum_{i=1}^{q} \theta_j^2}{2\sigma_\theta^2} - \frac{\sum_{k=1}^{m} \beta_k^2}{2\sigma_\beta^2} \Big]. \tag{18}$$

A simpler alternative, obtained as the limit of the prior distribution in (18), is the uniform prior, given by $\pi(\delta) = c \in \Re$ for all possible values of δ, and 0 otherwise.

Prediction of future values can be obtained through the predictive distribution, given by

$$p(y_{n+h} \mid Y^{(n)}) = \int p(y_{n+h} \mid Y^{(n)}, \delta)\pi(\delta \mid Y^{(n)})d\delta.$$

Once a sample $\delta^{(1)}, ..., \delta^{(M)}$ is available, the h-step-ahead predictive distribution can be approximated by

$$p(y_{n+h} \mid Y^{(n)}) \cong \frac{1}{M}\sum_{i=1}^{M} p(y_{n+h} \mid Y^{(n)}, \delta^{(i)}).$$

As the posterior distribution does not have a closed form, Markov Chain Monte Carlo (MCMC) methods can be employed to estimate the parameter vector. More specifically, a Metropolis-Hastings (M-H) algorithm, a powerful MCMC method, is highly advised when simulating multivariate distributions (see Gamerman and Lopes (2006) for details).

In the M-H algorithm, it is necessary to specify a candidate kernel. An usual choice is based on a random walk chain, with a multivariate normal density,

$$\delta^{(r)} = \delta^{(r-1)} + \omega_r$$

where $\omega_r \sim N(0, cH)$, H is given by the Hessian matrix of the maximum likelihood estimates and c is the tuning parameter, designed to optimize the convergence of the algorithm.

In this procedure, all the components of parameter vector δ are updated in a single block. Besides, the use of the hessian matrix enables the algorithm to walk in the right direction in a faster way. The definition of the tuning parameter is also an important feature to attain, for instance, a reasonable acceptance rate. If ω_r is large, the acceptance rate can be small, while if ω_r is too small, it will be necessary many steps to reach convergence.

The convergence of the algorithm can be checked through the Gelman and Rubin method Gelman (1966), in which multiple chains with different initial values are used. It is also possible to analyze the trajectory (trace) and histogram plots of the generated chain, as suggested by Gamerman and Lopes (2006).

4 Empirical results

An extensive Monte Carlo experiment was conducted to compare the performance of the observation and parameter driven approaches using the Poisson model. The generation and estimation steps were performed in the 3.0.2 R

package. The Zeger-ARp model, in this section abbreviated to ZEG, was fitted using R-INLA and the GARMA and GLARMA models were fitted using MCMC algorithms (available upon request).

The series were generated under each one of the models described in Section 2, and the three models were fitted to all series. Thus we have the following scenarios for the data generating process (DGP) using a conditional Poisson distribution for y_t:

Scenario 1: (ZEG) The DGP is the Zeger model.

Scenario 2: (GARMA) The DGP is the GARMA model.

Scenario 3: (GLARMA) The DGP is the GLARMA model.

The next subsections present the results considering two structures for the mean process, μ_t: a pure autoregressive or an autoregressive plus a covariate.

4.1 Model with a pure autoregressive in the mean function

Series were generated for the 3 scenarios with

$$\mu_t = \exp(Z_t)$$

where Z_t in each model is given by

ZEG: $Z_t = \phi Z_{t-1} + u_t, \quad u_t \sim i.i.d.N(0,1)$,

GARMA: $Z_t = \phi(ln(y^*_{t-1}))$, where y^*_{t-j}=max$(y_{t-j}, 0.1)$.

GLARMA: $Z_t = \phi(Z_{t-1} + e_{t-1})$, where $e_t = (y_t - \mu_t)\mu_t^{-\lambda}$ and $\lambda = 0.5$.

Three values were chosen for $\phi = (0.4, 0.7, 0.95)$, with the last one very close to the nonstationary region.

We have initially performed a simulation to study the behavior of the methods for large series. Thus, one series of size n=10,000 was generated with a pure autoregressive and the results are presented in Table 1. To give an idea of the ability of each model to estimate the generated mean function, the table also presents the mean square error of the adjusted mean (MSE($\hat{\mu}$)), calculated as

$$MSE(\hat{\mu}^{(proc)}) = \frac{1}{n}\sum_{t=1}^{n}(\hat{\mu}_t^{(proc)} - \mu_t^{(proc)})^2,$$

where *proc* = ZEG, GARMA or GLARMA procedures.

Table 1: $\hat{\phi}$ and MSE($\hat{\mu}$) for models generated from an AR(1) process in the mean (n=10,000).

DGP	Fitted Model	Real value of parameter ϕ					
		$\phi = 0.4$		$\phi = 0.7$		$\phi = 0.95$	
		$\hat{\phi}$	MSE($\hat{\mu}$)	$\hat{\phi}$	MSE($\hat{\mu}$)	$\hat{\phi}$	MSE($\hat{\mu}$)
ZEG	ZEG	**0.412**	**1.314**	**0.703**	**2.346**	**0.949**	**219.92**
	GARMA	0.241	7.453	0.776	28.68	0.996	$2.1{\cdot}10^{+6}$
	GLARMA	0.096	9.115	0.038	93.12	0.000	$6.6{\cdot}10^{+6}$
GARMA	ZEG	0.764	0.129	0.849	0.252	0.944	0.718
	GARMA	**0.399**	**0.00004**	**0.705**	**0.00009**	**0.956**	**0.001**
	GLARMA	0.403	0.096	0.595	0.451	0.250	2.288
GLARMA	ZEG	0.739	0.264	0.847	1.416	**0.975**	**89.52**
	GARMA	0.354	0.183	0.848	3.040	0.995	$5.4{\cdot}10^{+4}$
	GLARMA	**0.406**	**0.0002**	**0.700**	**0.003**	0.001	$1.3{\cdot}10^{+5}$

Obs.: In bold are the estimates closer to the real values and the smallest MSE($\hat{\mu}$).

As expected, the estimates closer to the real value and the smallest MSE($\hat{\mu}$) are obtained when the fitted model is the same as the DGP. The only exception are the results for the series generated under the GLARMA process with $\phi = 0.95$, where the ZEG procedure provided better results. In fact, when ϕ is close to the nonstationary region, the ZEG seems to be the more stable procedure for estimation, regardless the DGP used, as it gives estimates closer to 0.95 and reasonable fits. For other values of ϕ there is not one approach that is better in all situations.

The GLARMA model undergo many problems, as we can notice from Table 1. When the DGP is the ZEG procedure, the GLARMA model gives estimates of ϕ close to zero for all cases and, when the DGP is GARMA, it only gives good results for $\phi = 0.4$. In addition, the GLARMA procedure presents serious problems even when the data are generated by this same process, unless $\phi = 0.4$. It can be perceived that, when $\phi = 0.95$, the estimate is close to zero, with a large MSE($\hat{\mu}$). When $\phi = 0.7$, the GLARMA approach seems to perform very well, but a close inspection reveals that the log-likelihood for the generated series under this process presents a very peculiar behavior, as it can be seen from Figure 1, which also presents the log-likelihood for the GARMA model in this case. The log-likelihood function shows a local maximum around 0.06 and a global maximum in 0.7, but with very few points around this value. Thus, it is very difficult to reach the global maximum, unless the initial value in the optimization algorithm

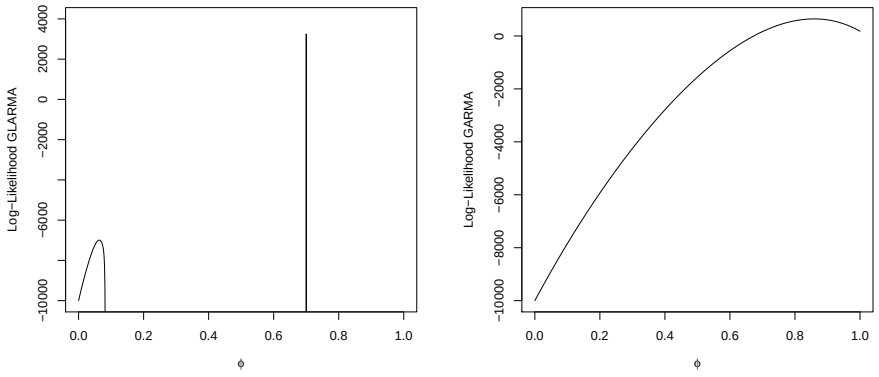


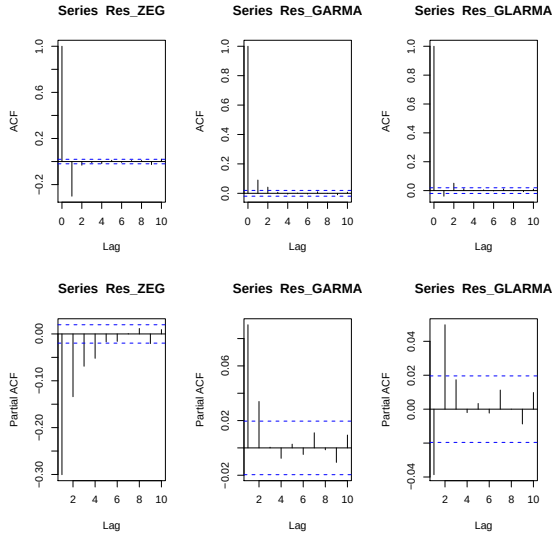
Figure 1: Likelihood for the GARMA and GLARMA models when the DGP is a GLARMA with $\phi = 0.7$ (n=10,000).

is very close to the real value. This seems to happen very often in this procedure, as different series generated under this process have repeated this same behavior.

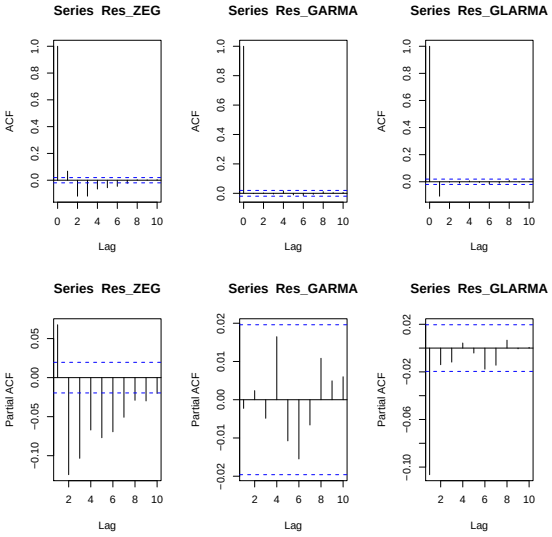
A detailed analysis on the behavior of the generated series and the fitted models reveals some interesting characteristics (results not presented here, but available upon request). Concerning the generation process, the ZEG DGP produces series with a few very large values. This is the reason why the models fitted to the series generated by the ZEG procedure possess MSE($\hat{\mu}$) larger than the other DGP's, as it was observed in Table 1. Another comment regarding the generation process is that, although all the series were generated by an AR(1) process in the mean, the GARMA DGP is the only approach which results in an AR(1) process in the data, y_t.

In relation to model fitting, residual analysis does not show, in general, any inadequacies concerning the assumptions of zero mean and homoscedasticity. The residuals seem to present a random pattern around zero, although there are some outliers in the GARMA and GLARMA fits when the DGP is the ZEG process. The problem, however, arises in the independence assumption. The GARMA and GLARMA models, when applied to the series generated by their respective DGP's, return white noise residuals. Nevertheless, the ZEG model always produces autocorrelated residuals (see Figure 2). This fact was already noticed by Zeger (1988), which observed a statistically significant autocorrelation in the residuals of the model fitted to the real data of polio infection in the U.S.

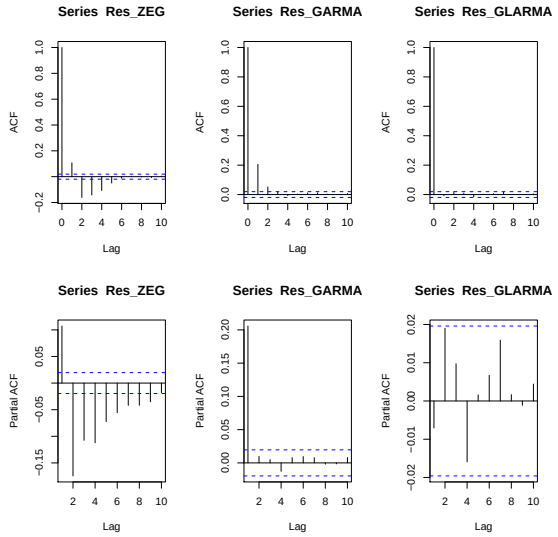
Table 2 shows the results for small series (n=100). Values in the table are



(a) DGP: ZEG process



(b) DGP: GARMA process



(c) DGP: GLARMA process

Figure 2: Autocorrelation (ACF) and partial autocorrelation (PACF) functions for residual series obtained by fitting the ZEG, GARMA and GLARMA models with $\phi = 0.4$ (n=10,000)

the mean of the estimates, obtained from a Monte Carlo (MC) experiment with 1000 replications. The behavior is very similar to the one presented by large series (Table 1). The main difference concerns the results for the ZEG model fitted to the GARMA and GLARMA DGP's when $\phi = 0.4$. While this procedure tends to overestimate the real ϕ for large series, it is not able to identify the AR(1) component for small series. Regarding the GLARMA estimates for the GLARMA DGP, we can observe once more the difficulty of this procedure in estimating values of ϕ that approximate the nonstationary region. It gives very poor estimates for $\phi = 0.95$ and, for $\phi = 0.7$, although the mean value is not far ($\hat{\phi} = 0.683$), its MSE (MSE($\hat{\phi}$)) is larger than the other procedures.

Table 2: Mean of $\hat{\phi}$ and MSE($\hat{\mu}$) in 1000 MC replications, for models generated from an AR(1) process in the mean (n=100).

DGP	Fitted Model	Real value of parameter ϕ					
		$\phi = 0.4$		$\phi = 0.7$		$\phi = 0.95$	
		$\hat{\phi}$	MSE($\hat{\mu}$)	$\hat{\phi}$	MSE($\hat{\mu}$)	$\hat{\phi}$	MSE($\hat{\mu}$)
ZEG	ZEG	**0.398**	**1.235**	**0.698**	**1.871**	**0.916**	**11.779**
		(0.029)		**(0.013)**		**(0.022)**	
	GARMA	0.195	5.592	**0.699**	13.57	0.973	948.89
		(0.073)		(0.045)		(103.08)	
	GLARMA	0.128	5.505	0.152	15.29	0.043	1536.33
		(0.074)		(0.295)		(0.754)	
GARMA	ZEG	0.001	0.166	0.783	0.302	0.915	0.494
		(0.157)		(0.211)		(0.139)	
	GARMA	**0.397**	**0.006**	**0.693**	**0.010**	**0.933**	**0.018**
		(0.008)		**(0.009)**		**(0.008)**	
	GLARMA	0.409	0.085	0.638	0.225	0.641	0.594
		(0.012)		(0.018)		(0.137)	
GLARMA	ZEG	0.001	0.285	0.840	1.261	**0.952**	**18.23**
		(0.144)		(0.025)		**(0.012)**	
	GARMA	0.336	0.135	0.836	1.627	0.984	830.50
		(0.017)		**(0.023)**		(0.028)	
	GLARMA	**0.398**	**0.013**	**0.683**	**0.079**	0.049	1527.39
		(0.006)		(0.043)		(0.696)	

Obs.: In bold are the estimates closer to the real values and the smallest MSE($\hat{\mu}$). In parentheses are the MSE of $\hat{\phi}$.

4.2 Models with a covariate plus autoregressive terms in the mean function

Monte Carlo experiments with 1000 replications for series of size n=100 were performed for the three scenarios with

$$\mu_t = \exp(\alpha + \beta x_t + Z_t).$$

We allowed the covariate x_t to take one of the following forms: $x_t \sim N(0,1)$ or x_t is an AR(1) process with $\phi = 0.4$. The component Z_t in each model is given by

ZEG: $Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + u_t, \quad u_t \sim i.i.d.N(0,1)$,

GARMA: $Z_t = \sum_{i=1}^{p} \phi_i \big(ln(y^*_{t-i}) - \alpha - \beta x_{t-i}\big)$, where y_{t-i}=max$(y_{t-i}, 0.1)$.

GLARMA: $Z_t = \sum_{i=1}^{p} \phi_i (Z_{t-i} + e_{t-i})$, where $e_t = (y_t - \mu_t)\mu_t^{-\lambda}$.

Two sets of simulations were performed in this case. The first one with AR(1) components in the mean process and the second one with AR(2) components.

CASE 1: AR(1) ($p = 1$)

In this case, $\phi = 0.4$ or 0.7, $\alpha = 1.0$, $\beta = 0.3$ and $\lambda = 0.5$. We also provide the MSE($\hat{\mu}$) and prediction mean square errors (PMSE) for forecasts 1, 5 and 15 steps-ahead, to compare the procedures in all scenarios. The GLARMA DGP when $\phi = 0.95$ presented many problems, with series composed entirely of zero values, or series with outliers of a very large magnitude. Thus, it was not possible to perform the MC replications in this case.

In what follows, only the results for the covariate $x_t \sim N(0,1)$ are presented. The conclusions were very similar when the covariate is a time series ($x_t \sim AR(1)$ process).

Figure 3 presents boxplots for the estimation of α, β and ϕ under the three DGP's, the covariate $x_t \sim N(0,1)$ and the two values considered for ϕ in the AR(1) process. The first conclusion we can draw from the figure concerns the estimation of α, which seems to be best estimated by the method which generated the data, as expected. We can also notice that the GARMA and GLARMA procedures overestimate this parameter when the true DGP is the Zeger method, while α is underestimated by the ZEG and GLARMA procedures when the true DGP is GARMA. For the GLARMA DGP, only the GARMA methodology seems to overestimate α.

Analysing the estimation of the coefficient of the covariate, β, it is interesting to note that all the procedures are robust with respect to the estimation of this parameter, regardless the magnitude of the autocorrelation structure introduced in the mean function. This fact has already been noticed by Jung et al. (2006) and Jung and Tremayne (2011) when applying similar parameter and observation driven models to some real data set. Nevertheless, it should be pointed out that the variability is larger when the data are generated by the ZEG model, and the GLARMA procedure produces the largest variability in this case. When $\phi = 0.7$ and data are generated from the GLARMA DGP, the ZEG and GARMA estimates of β seems to overestimate the real β. When the observations are generated from the GARMA DGP with $\phi = 0.4$ or 0.7, the GLARMA underestimates the real β.

Estimates of ϕ are shown at the bottom of Figure 3. It is clear from this figure that the best estimates are obtained by the same DGP from which the data were generated. Analyzing the cases ϕ=0.4 and 0.7, we see that the GARMA approach always underestimates the real ϕ when the observations are generated from other DGP's, while the ZEG procedure tends to overestimate ϕ from other DGP's. The GLARMA, in general, underestimates the real ϕ, unless $\phi = 0.4$ and observations are generated by the GARMA or the own GLARMA procedure.

Table 3 shows mean estimates, in 1000 MC replications, of MSE($\hat{\mu}$) and PMSE for forecasts h =1, 5 and 15 steps-ahead for each procedure, still considering the covariate $x_t \sim N(0, 1)$. Similar to the case of no covariates (Section 4.1) the smallest MSE($\hat{\mu}$)'s are obtained when the estimation procedure is the same as the DGP. Concerning the predictions, the GLARMA approach presents a very good performance when data are generated by the ZEG and GLARMA DGP's, and it still gives some reasonable results when the DGP is the GARMA procedure.

CASE 2: AR(2) ($p = 2$)

For the AR(2) case, we fixed $(\phi_1, \phi_2) = (0.25, 0.25)$ or (0.5, 0.25) and α and β again equal to 1 and 0.3, respectively. As observed in Section 3.2, parameter estimation in the GLARMA procedure is very sensitive to the value of λ. It seems that, in the AR(2) case, $\lambda = 1$ is a better choice compared to $\lambda = 0.5$, as the estimates are closer to the real values, and the likelihood is more well behaved. Thus in this section we have used $\lambda = 1$ in the simulations. Once again, only the results for covariate $x_t \sim N(0, 1)$ are presented.

Table 4 presents the mean values, over MC=1000 replications for the estimates, with MSE, of parameters α, β, ϕ_1 and ϕ_2. The main features

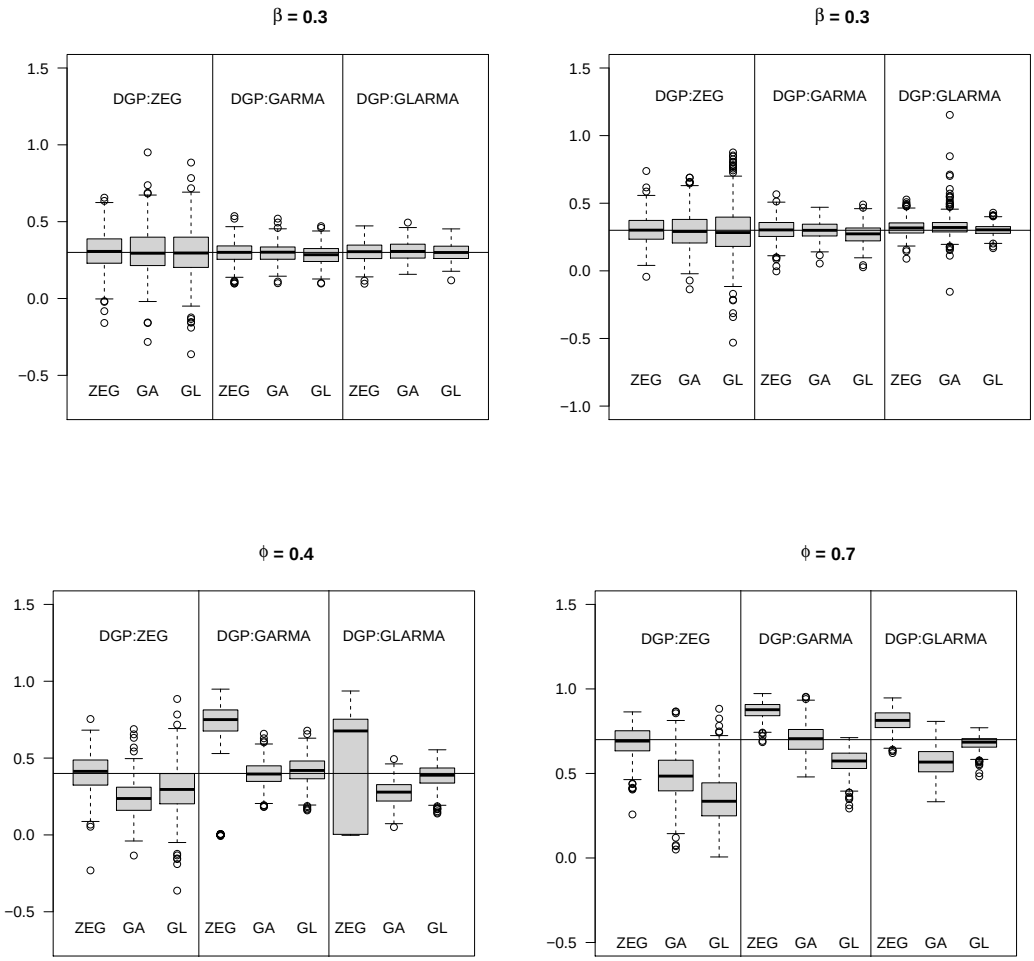


Figure 3: Boxplot for the estimates of α, β and ϕ_1, in 1000 MC replications, for series of size $n = 100$. GA means GARMA and GL means GLARMA. A straight horizontal line is placed on the real value of the parameters.

Table 3: Mean of MSE($\hat{\mu}$) and PMSE in 1000 MC replications (n=100).

	DGP	Fitted Model	MSE($\hat{\mu}$)	PMSE		
				$h=1$	$h=5$	$h=15$
$\phi=0.4$	ZEG	ZEG	**4.436**	7.500	12.292	11.251
		GARMA	39.043	7.228	16.313	13.653
		GLARMA	39.790	**6.131**	**10.378**	**9.345**
	GARMA	ZEG	0.626	1.409	1.698	1.399
		GARMA	**0.080**	1.110	1.681	**1.334**
		GLARMA	0.277	1.501	**1.588**	1.453
	GLARMA	ZEG	0.627	1.694	**1.456**	1.994
		GARMA	0.227	1.555	1.490	2.029
		GLARMA	**0.098**	**1.475**	**1.457**	**1.859**
	DGP	Fitted Model	MSE($\hat{\mu}$)	PMSE		
				$h=1$	$h=5$	$h=15$
$\phi=0.7$	ZEG	ZEG	**5.594**	**8.881**	32.184	31.690
		GARMA	87.087	**8.874**	51.772	52.766
		GLARMA	95.032	11.283	**21.437**	**18.143**
	GARMA	ZEG	0.858	0.770	**1.637**	2.216
		GARMA	**0.093**	**0.427**	1.680	2.881
		GLARMA	0.503	0.612	2.011	**2.031**
	GLARMA	ZEG	1.762	1.690	3.629	3.717
		GARMA	0.998	1.452	3.384	**3.534**
		GLARMA	**0.225**	**1.223**	**3.161**	3.742

Obs.: In bold are the smallest MSE($\hat{\mu}$) and PMSE.

observed in the AR(1) case are also perceived here. The first conclusion is that the best model to estimate the parameters is the one that generated the data. The second distinctive characteristic is that the procedures are robust to the estimation of β, the covariate parameter. That is, regardless the DGP used, the β estimates are always close to the real value of this parameter, for all fitted models. The main differences among the methods occur generally in the intercept parameter, α, especially for the combination of the autoregressive parameters closer to the non-stationary region ($\phi_1 = 0.50, \phi_2 = 0.25$). The parameter driven approach (ZEG) also presents some difficulties in estimating the autoregressive parameters, ϕ_1 and ϕ_2, when the DGP is GARMA or GLARMA.

Table 4: Mean of $\hat{\alpha}$, $\hat{\beta}$, $\hat{\phi_1}$ and $\hat{\phi_2}$ in 1000 MC replications, for models generated from an AR(2) process in the mean (n=100).

		$\phi_1 = 0.25$, $\phi_2 = 0.25$				$\phi_1 = 0.50$, $\phi_2 = 0.25$			
DGP	Fitted Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\phi_1}$	$\hat{\phi_2}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\phi_1}$	$\hat{\phi_2}$
	ZEG	**1.009**	**0.295**	**0.246**	**0.226**	**1.010**	**0.300**	**0.493**	**0.218**
		(0.043)	(0.014)	(0.018)	(0.020)	(0.151)	(0.014)	(0.022)	(0.024)
Z	GARMA	0.294	**0.295**	0.158	0.147	-22.705	0.298	0.346	0.185
E		(3455)	(0.019)	(0.019)	(0.021)	(57268)	(0.024)	(0.044)	(0.115)
G	GLARMA	1.446	0.291	0.148	0.146	1.593	0.291	0.254	0.211
		(0.304)	(0.024)	(0.021)	(0.023)	(0.617)	(0.032)	(0.071)	(0.095)
G	ZEG	0.353	0.302	1.420	-0.536	0.353	**0.302**	0.956	-0.158
A		(1.061)	(0.011)	(097)	(0.347)	(0.149)	(0.010)	(0.573)	(0.275)
R	GARMA	**0.966**	**0.301**	**0.253**	**0.236**	**0.819**	0.307	**0.477**	0.229
M		(0.030)	(0.005)	(0.006)	(0.006)	(0.396)	(0.007)	(0.010)	(0.082)
A	GLARMA	0.601	**0.299**	0.230	0.222	-0.013	0.308	0.385	**0.250**
		(0.211)	(0.005)	(0.006)	(0.006)	(1.281)	(0.010)	(0.018)	(0.069)
G	ZEG	0.962	0.311	1.420	-0.536	0.980	**0.303**	0.935	-0.280
L		(0.090)	(0.003)	(0.886)	(0.652)	(0.017)	(0.004)	(0.822)	(0.419)
A	GARMA	1.148	0.306	0.169	0.160	1.411	0.314	0.406	0.143
R		(0.037)	(0.003)	(0.012)	(0.014)	(0.207)	(0.003)	(0.027)	(0.140)
M	GLARMA	**0.982**	**0.300**	**0.240**	**0.234**	**1.010**	**0.303**	**0.464**	**0.238**
A		(0.018)	(0.003)	(0.008)	(0.008)	(0.085)	(0.003)	(0.009)	(0.077)

Obs.: In bold are the estimates closer to the real values. In parentheses are the MSEs.

5 Conclusions

This study was motivated by the increasing use of generalized linear models with autoregressive components in the mean to model the autocorrelation present in time series of counts. Although there are some works devoted to establish the theoretical properties of either observation and parameter driven models, and some other works which aim to compare these methodologies using real data set, we believe this to be the first study which intends to compare the two approaches based on a Monte Carlo experiment.

The observation driven models addressed here are the GARMA and GLARMA procedures, proposed respectively by Benjamim et al. (2003) and Davis et al. (2003). In their work, the parameter estimation is performed using the maximum likelihood method, while in this paper we propose a Bayesian approach based on MCMC. The parameter driven model used here, originally proposed by Zeger (1988), was fitted using the R-INLA package.

The extensive simulation study was performed in models with and with-

out a covariate. The main conclusion that can be draw is that the parameter of the covariate, β, is not affected by the model chosen to fit the data, regardless the proccess which generate the data. The only exception occurs in the case where the covariate is a function of time, $x_t = t/n$, when the magnitude of the autocorrelation can introduce some bias in the estimation of β. This can be an indication that the methods are robust with respect to the estimation of β. The same inference can not be taken with respect to the autoregressive and intercept, as they seem to be well estimated only using the same model that generates the data.

Concerning the other parameters, the one related with the intercept, α, and the parameters associated with the autocorrelation structure, ϕ, we have observed that they are well estimated only by the own model which generated the series.

Some other remarks concern the estimation of the GLARMA model, which is highly sensitive to the parameter λ used to standardize the residuals. For AR(1) models it seems that the best value for λ is close to 0.5, while for AR(2) models it is better to use a larger value (close to 1). When the autoregressive parameter is very close to the non-stationary region (for example $\phi = 0.95$ in the AR(1)) the GLARMA model presents the worst performance, most of the times not returning a valid estimate for the parameters.

With respect to the parameter driven model, we have observed some problems regarding the residual series. After properly estimating the parameters, the residuals, which should be a white noise, still presents an autocorrelation structure.

Acknowledgements

G. C. Franco gratefully acknowledges FAPERJ, FAPEMIG and CNPq/Brazil for partial financial support. H. S. Migon gratefully acknowledges FAPERJ and CNPq/Brazil and M. O. Prates acknowledges FAPEMIG for partial financial support.

References

Benjamin, M. A., R. A. Rigby, and D. M. Stasinopoulos (2003). Generalized autoregressive mov-ing average models. *Journal of the American Statistical Association*, **98**, pp. 214-223.

Box, G.E.P. and Jenkins, G.M. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Chan, K. and J. Ledolter (1995). Monte Carlo EM Estimation for Time Series Models Involving Counts. *Journal of the American Statistical Association*, **90**, pp 242-251.

Cox, D.R. (1981). Statistical Analysis of Time Series: Some Recent Developments. *Scandinavian Journal of Statistics*, **8**, pp. 93-115.

Creal, D., Koopman, S. J. and Lucas, A. (2013). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*, **28**, pp 777-795.

Czado, C. and Kolbe, A. (2004) Empirical Study of Intraday Option Price Changes using extended Count Regression Models. Collaborative Research Center 386, Discussion Paper 403. Ludwig-Maximilians-Universitat, Germany.

Davis, R.A., Dunsmuir, W.T.M. and Wang, Y. (1999). Modelling time series of counts. In: Ghosh, S. (Ed.), Asymptotics, Nonparametrics, and Time Series: a Tribute to Madan Lal Puri. Marcel Dekker, NewYork, pp. 63-113.

Davis, R.A., Dunsmuir, W.T.M. and Streett, S.B. (2003). Observation-driven models for Poisson counts. *Biometrika*, **90**, pp 777-790.

Davis, R.A., Dunsmuir, W.T.M. and Streett, S.B.: Maximum likelihood estimation for an observation driven model for Poisson counts. Methodol. *Comput. Appl. Probab.*, **7**, 149-159.

Diebold, F.X. and Schuermann, T. (1996). Exact Maximum Likelihood Estimation of Observation-Driven Econometric Models. *NBER Technical Working Paper No. 194.*

Durbin, J., Koopman, S.J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, **84**, pp 669-684.

Durbin, J., Koopman, S.J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *J. Roy. Statist. Soc. Ser. B*, **62**, pp 3-56.

Feigin, P.D., Gould, P., Martin, G.M. and Snyder, R.D. (2008) Feasible parameter regions for alternative discrete state space models. *Stat. Probab. Lett.*, **78**, 2963-2970.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall.

Gamerman, D., Santos, T. R. and Franco, G. C. (2013). A non-Gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, **34**, pp 625-645.

Gelman, A. (1996). *Inference and monitoring convergence*. In Markov Chain Monte Carlo in Practice. W.R. Gilks, Richardson, S. Spiegelhalter, D. J. (eds). London: Chapman and Hall, pp. 131-143.

Harvey, A.C. and Fernandez, C. (1989). Time series models for count or qualitative observations. *J. Bus. Econom. Statist*, **7**, pp 407-417.

Jung, R. C., Kukuk, M. and Liesenfeld, R. (2006). Time series of count data: modeling, estimation and diagnostics. *Computational Statistics and Data Analysis*, **51**, pp. 2350-2364.

Jung, R. C. and Tremayne, A. R. (2011). Useful models for time series of counts or simply wrong ones? *Advances in Statistical Analysis*, **95**, pp 59-91.

Kedem, B. and Fokianos, K. (2002) *Regression Models for Time Series Analysis*. New Jersey: John Wiley & Sons.

Li, W. K. (1994). Time Series Models Based on Generalized Linear Models: Some Further Results. *Biometrics*, **50**, pp 506-511.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, pp 319-392.

Rydberg, T.H. and Shephard, N. (1999). BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time. *Working paper, Oxford University*.

Shephard, N. (1995). Generalized linear autorgressions. *Working Paper, Nuffield College, Oxford.*.

Shephard, N., Pitt, M.K. (1995). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, **84**, pp 653-667.

Smith, R. L. and Miller, J. E. (1986). A non-Gaussian state space model and application to prediction of records. *J. R. Statist. Soc. B*, **48**, pp 79-88.

Tierney, L. and Kadane, J. B., (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the Americal Statistical Association*, *bf 81*, pp 82-86.

West, M., Harrison, J. and Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting. *Journal of the Americal Statistical Association*, *bf 80*, pp 73-83.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, **75**, pp 621-629.

Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, **44**, pp 1019-1031.