Brazilian Journal of Probability and Statistics
2019, Vol. 33, No. 4, 756–781
https://doi.org/10.1214/19-BJPS437

Time series of count data: A review, empirical comparisons and data analysis

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

[a]*Department of Statistics, Universidade Federal de Minas Gerais*
[b]*Department of Statistical Methods, Universidade Federal do Rio de Janeiro*

Abstract. Observation and parameter driven models are commonly used in the literature to analyse time series of counts. In this paper, we study the characteristics of a variety of models and point out the main differences and similarities among these procedures, concerning parameter estimation, model fitting and forecasting. Alternatively to the literature, all inference was performed under the Bayesian paradigm. The models are fitted with a latent AR(p) process in the mean, which accounts for autocorrelation in the data. An extensive simulation study shows that the estimates for the covariate parameters are remarkably similar across the different models. However, estimates for autoregressive coefficients and forecasts of future values depend heavily on the underlying process which generates the data. A real data set of bankruptcy in the United States is also analysed.

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 and can be illustrated by means of generalized state-space models, as pointed out in Davis, Dunsmuir and Wang (1999). Using this formulation, the observation equation remains the same for the two approaches, while the state equations are distinctly defined. In the observation driven case, the distribution of the state equation is specified conditional on past observations, while in the parameter driven approach, it is conditioned on a latent dynamic process.

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, Benjamin, Rigby and Stasinopoulos (2003) compiled the results of these two works in a more

Key words and phrases. Observation driven model, parameter driven model, autoregressive processes, Bayesian inference.

Received May 2018; accepted March 2019.

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, Dunsmuir and Wang (1999), Davis, Dunsmuir and Streett (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, lately proposed by Creal, Koopman and Lucas (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. An empirical comparison of the relative forecasting performance of the GAS model with respect to some parameter and observation driven models was performed by (Koopman, Lucas and Scharth, 2016).

Concerning parameter driven models, in the Bayesian context there is the seminal paper of West, Harrison and Migon (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, Santos and Franco (2013) and Souza and Migon (2018). Estimation of non-Gaussian state space models can also be performed using sequential Monte Carlo methods (particle filters), see Andrieu, Doucet and Holenstein (2010).

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, Kukuk and Liesenfeld (2006) and Czado and Kolbe (2004). State-space models with non-Gaussian observations were also considered in a few works For example, Shephard and Pitt (1995) have used MCMC and importance sampling to perform, respectively, Bayesian and classical analysis. Durbin and Koopman (2002) have presented a new, simple and computationally efficient simulation smoother, employing an approach in which only mean

corrections for unconditional vectors are required. Creal (2012) has provided an extensive survey on the use of sequential Monte Carlo methods in non-Gaussian state space models. Migon et al. (2005) have proposed an alternative way to obtain a sample from the states (direct multivariate normal sampling) by direct evaluation of the prior full conditional of the full state vector, combined with the likelihood in matrix form.

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, few 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, Dunsmuir and Wang (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, Kukuk and Liesenfeld (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.

One of the main contributions of the paper is towards understanding the distinction between the observation and parameter driven models for time series of counts. In order to achieve this, a large and detailed simulation study is presented to investigate the similarities and differences between the two approaches and studied models. 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, Kukuk and Liesenfeld (2006), efficient estimation of the Zeger (1988) model is difficult to perform, as the likelihood function depends on high-dimensional integrals. The WinBUGS project (Lunn et al. (2000)), or the parallel version MultiBUGS (Goudie et al. (2017)), are a flexible alternative to perform Bayesian inference and may be used to fit parameter driven models, see for example the paper of Meyer and Yu (2000) with an application to stochastic volatility models. Nevertheless, in this work we propose to use the integrated nested Laplace approximation (INLA) approach (Rue, Martino and Chopin (2009)), which provides efficient Bayesian inference for the marginal posterior densities of the hyperparameters and the latent components. Concerning observation driven models, as far as the authors know, inference for the GARMA and GLARMA models was only performed under the classical perspective. 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 INLA and MCMC algorithms. Section 4 discusses the results of the Monte Carlo experiments. Section 5 presents a real application on bankruptcy data in the US. Finally, some conclusions are drawn in Section 6.

2 Models

This section presents the parameter and observation driven models discussed in this paper. 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 $\{y_t\}_{t=1,\ldots,n}$ be a non-Gaussian time series, with conditional distribution belonging to the exponential family,

$$p(y_t|\mathscr{H}_t) = a(y_t, \varphi)\exp\left[\frac{1}{\varphi}(y_t\vartheta_t - b(\vartheta_t))\right], \tag{1}$$

where ϑ_t and φ are the canonical and scale parameters and $a(\cdot)$ and $b(\cdot)$ are specific functions. $\mathscr{H}_t$ is a σ-field generated either by past values of the observed series (observation driven models) or by a latent process (parameter driven models), as well as past and present values of the covariates.

The conditional mean, $\mu_t = E(y_t \mid \mathscr{H}_t)$, is related to the predictor, η_t, by a link function, g,

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

where $\boldsymbol{x}_t = (x_1, \ldots, x_m)'$ is a $m \times 1$ vector of covariates, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$ and the structure of Z_t, which accounts for the autocorrelation present in the process, depends on the adopted approach, the parameter or the observation driven model. Some sub-models for count series include, for example, Poisson, binomial and negative binomial distributions.

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 predictions, 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, Dunsmuir and Wang (1999)).

In the next subsections, we present the procedures in more detail, focusing on the Poisson and negative binomial distributions, which enable a more straightforward comparison among the models.

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 non-negative weakly stationary latent process, ϵ_t, and on $\boldsymbol{x}_t$, the mean of y_t is given by

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

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

$$\lambda_t = E(y_t) = E\big(E\big(y_t \mid (\epsilon_t, \boldsymbol{x}_t)\big)\big) = \exp(\boldsymbol{x}_t'\boldsymbol{\beta})$$

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.

Taking the special cases of the Poisson and negative binomial and assuming that $\sigma^2 = \text{var}(\epsilon_t)$, the marginal variance and the autocorrelation function of y_t are given by,

Poisson:

- $\text{var}(y_t) = E(\text{var}(y_t \mid (\epsilon_t, \boldsymbol{x}_t))) + \text{var}(E(y_t \mid (\epsilon_t, \boldsymbol{x}_t))) = \lambda_t + \lambda_t^2\sigma^2$;
- $\rho_y(t,s) = \text{corr}(y_t, y_s) = \frac{\rho_\epsilon(s-t)}{[(1+(\sigma^2\lambda_t)^{-1})(1+(\sigma^2\lambda_{t+s})^{-1})]^{1/2}}$.

Negative Binomial:

- $\text{var}(y_t) = \lambda_t + \lambda_t^2((r+1)\sigma^2 + 1)/r$;
- $\rho_y(t,s) = \text{corr}(y_t, y_s) = \rho_\epsilon(s-t)/[((r+1)/r + (r\sigma^2)^{-1} + (\sigma^2\lambda_t)^{-1})((r+1)/r + (r\sigma^2)^{-1} + (\sigma^2\lambda_{t+s})^{-1})]^{1/2}$;

where r is the precision parameter and $\rho_\epsilon(s-t)$ is the correlation function between ϵ_t and ϵ_s.

In both cases, the degree of overdispersion depends on λ_t and the autocorrelation in y_t should be less than or equal that in ϵ_t, since $\lambda_t > 0$.

One possibility to model the latent process, ϵ_t, is to to assume that $Z_t = \ln(\epsilon_t)$ in (2) follows a stationary Gaussian autoregressive process of order p, in order to

ensure that the conditional mean of y_t is non-negative. As the logarithmic is the used link function for Poisson and negative binomial distributions, they are the natural candidates to be used in this case.

Thus, the parameter driven model stated here, called from now on Zeger-ARp, is given by

$$y_t \mid (\epsilon_t, \boldsymbol{x}_t) \sim G(\mu_t), \quad t = 1, \ldots, n, \tag{3}$$

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

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

where G can be the Poisson or negative binomial distributions, $\Phi(B) = (1 - \sum_{j=1}^{p} \phi_j B^j)$, 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, Kukuk and Liesenfeld (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 Benjamin, Rigby and Stasinopoulos (2003) and by Davis, Dunsmuir and Streett (2003). The similarities and differences between the two approaches are described in the next subsections in more detail.

2.2.1 *Generalized linear autoregressive moving average (GLARMA) model.* Davis, Dunsmuir and Streett (2003) provide the basis for the GLARMA model whose conditional distribution, given the previous information set, belongs to the exponential family form in Equation (1). In this case, the state equation is composed of a linear model for the explanatory variables plus an error term with an infinite moving average structure,

$$\eta_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{\infty} \tau_i u_{t-i}, \tag{6}$$

where u_t are the standardized residuals, given by,

$$u_t = \frac{(y_t - \mu_t)}{(\text{var}(y_t))^{\lambda}}, \quad \lambda \geq 0.$$

Davis, Dunsmuir and Streett (2003) show that, under the initial conditions $u_s = 0$ and $y_s = 0$, for $s \leq 0$, the u_t form a martingale difference sequence with

zero mean and variance $E(h(\mu_t))$, where $h(\mu_t) = \mu_t^{(1-2\lambda)}$, for the Poisson model and $h(\mu_t) = \mu_t^{(1-2\lambda)}/\zeta$, with $0 < \zeta \leq 1$, for the negative binomial. In addition, from the martingale difference property, $\operatorname{cov}(u_t, u_s) = 0$, for $t \neq s$. Then, it follows that $\operatorname{cov}(\eta_t, \eta_{t+j}) = \sum_{i=1}^{\infty} \tau_i \tau_{i+j} E(h(\mu_t))$, which does not depend on time t if $\lambda = 0.5$, even if μ_t is not strictly stationary.

In analogy with the ARMA models of Box and Jenkins (1976), Davis, Dunsmuir and Streett (2003) show that it is possible to write the infinite moving average term in Equation (6) as an autoregressive-moving average filter, by letting

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

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$ are polynomials with all roots outside the unit circle.

Davis, Dunsmuir and Streett (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. The authors have also implemented the glarma package in R language, to estimate the parameters of the GLARMA model for the Poisson, binomial and negative binomial distributions.

Thus the GLARMA(p, q) model, for the Poisson and negative binomial, is given by,

$$y_t \mid \left(Y^{(t-1)}, \boldsymbol{x}_t\right) \sim G(\mu_t), \quad t = 1, \ldots, n, \tag{7}$$

$$\eta_t = \ln(\mu_t) = \boldsymbol{x}_t'\boldsymbol{\beta} + Z_t, \tag{8}$$

$$Z_t = \sum_{i=1}^{\infty} \tau_i u_{t-i} = \sum_{i=1}^{p} \phi_i (Z_{t-i} + u_{t-i}) + \sum_{i=1}^{q} \theta_i u_{t-i}, \tag{9}$$

$$u_t = \left(y_t - e^{\eta_t}\right) e^{-\lambda \eta_t}, \tag{10}$$

where $Y^{(t-1)} = (y_1, \ldots, y_{t-1})$ and G can be the Poisson or negative binomial.

2.2.2 *Generalized autoregressive moving average (GARMA) model.* Benjamin, Rigby and Stasinopoulos (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. Following their approach, the predictor is given by the following general form,

$$\eta_t = g(\mu_t) = \boldsymbol{x}_t'\boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j \left(g(y_{t-j}) - \boldsymbol{x}_{t-j}\boldsymbol{\beta}\right) + \sum_{j=1}^{q} \theta_j \left(g(y_{t-j}) - \eta_{t-j}\right). \tag{11}$$

The second component in the right hand side of (11) represents the autoregressive term and consists of a function of the past values of y_t and $\boldsymbol{x}_t$. The third 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, $\ln(y_t) - \eta_t$.

Benjamin, Rigby and Stasinopoulos (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.

Although the procedures proposed in Benjamin, Rigby and Stasinopoulos (2003) can accommodate any model in which the conditional distribution of y_t, given a previous information set, belongs to the exponential family in Equation (1), in this paper we present only the results for the Poisson and negative binomial GARMA models, to keep in line with the other procedures addressed here. Thus, in our specific case,

$$y_t \mid \left(Y^{(t-1)}, \mu^{(t-1)}, \boldsymbol{x}_t\right) \sim G(\mu_t), \quad t = 1, \ldots, n, \tag{12}$$

$$\eta_t = \ln(\mu_t) = \boldsymbol{x}_t'\boldsymbol{\beta} + Z_t, \tag{13}$$

$$Z_t = \sum_{j=1}^{p} \phi_j \left(\ln\left(y_{t-j}^*\right) - \boldsymbol{x}_{t-j}\boldsymbol{\beta}\right) + \sum_{j=1}^{q} \theta_j \left(\ln\left(y_{t-j}^*/\mu_{t-j}\right)\right), \tag{14}$$

where G can be the Poisson or negative binomial distributions, $\mu^{(t-1)} = (\mu_{t-1}, \ldots, \mu_1)$ and $y_{t-j}^* = \max(y_{t-j}, c)$ with $0 < c < 1$. This last 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.

3 Estimation and computational aspects

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 $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ separately. Given consistent estimates for $\boldsymbol{\phi}$ and σ_u^2, which can be achieved

by a method of moments, the $\boldsymbol{\beta}$'s are estimated using quasi-likelihood. Davis, Dunsmuir and Wang (1999) use Taylor expansions to build an approximate likelihood, which is maximized through numerical algorithms. Another proposal is given in Jung, Kukuk and Liesenfeld (2006), using efficient importance sampling to approximate the complete likelihood, under the classical and Bayesian framework.

In this work, we propose to use the INLA approach (Rue, Martino and Chopin (2009)) to obtain efficient estimates for the parameter vector $\boldsymbol{\psi} = (\boldsymbol{\beta}', \boldsymbol{\phi}', \sigma_u^2)$. The likelihood function, $L(\boldsymbol{\psi}; Y^{(n)}, \boldsymbol{Z}) = \prod_{t=1}^{n} f(y_t \mid Z_t) p(Z_t \mid \boldsymbol{\psi})$, can be approximated by

$$L_P(\boldsymbol{\psi}; Y^{(n)}, \boldsymbol{Z}) \propto \left|\boldsymbol{V}^{-1}\right|^{1/2} \exp\left\{\sum_{t=1}^{n} (\eta_t y_t - e^{\eta_t}) - \frac{1}{2}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}\right\}$$

for the Poisson model and

$$L_{NB}(\boldsymbol{\psi}; Y^{(n)}, \boldsymbol{Z}) \propto \left|\boldsymbol{V}^{-1}\right|^{1/2} \exp\left\{\sum_{t=1}^{n} (\eta_t y_t - \ln(r + e^{\eta_t})) - \frac{1}{2}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}\right\}$$

for the negative binomial model, where $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$, $\boldsymbol{V}$ is the covariance matrix of $\boldsymbol{Z}$ and r is the precision parameter.

The INLA approach relies on a hierarchical structure, where the response y_t is assumed independent, conditional on some Gaussian latent field $\boldsymbol{w}$ and a vector of hyperparameters $\boldsymbol{\varsigma} = (\boldsymbol{\phi}', \sigma_u^2)$. These models are called latent Gaussian models (LGM). The LGM class assumes that the observation (or response) variable has 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) and the Gaussian assumption is also satisfied in (5). Therefore, these characteristics clearly allow for the use of the INLA framework.

The INLA method approximates the posteriors of interest with a closed form expression, which provides a substantial gain in computational time if compared with MCMC methods. Moreover, INLA does not have problems of convergence and mixing, inherent to MCMC runs. In our specific framework, the latent field is composed by the linear predictor and the regression coefficients, $\boldsymbol{w}_t = (\eta_t, \boldsymbol{\beta}')$, and the AR coefficients and the noise variance define the hyperparameter vector $\boldsymbol{\varsigma} = (\boldsymbol{\phi}', \sigma_u^2)$.

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

$$\pi(w_{tj}|Y^{(n)}) = \int \pi(w_{tj}|Y^{(n)}, \boldsymbol{\varsigma}) \pi(\boldsymbol{\varsigma}|Y^{(n)})\, d\boldsymbol{\varsigma}, \tag{15}$$

$$\pi(\varsigma_k|Y^{(n)}) = \int \pi(\boldsymbol{\varsigma}|Y^{(n)})\, d\boldsymbol{\varsigma}_{-k}, \quad k = 1, \ldots, p+1, \tag{16}$$

where $\boldsymbol{\varsigma}_{-k}$ denotes vector $\boldsymbol{\varsigma}$ without the kth component and w_{tj} is the jth entry of vector $\boldsymbol{w}_t$.

The posterior marginals presented in (15) and (16) are calculated based on a Laplace approximation to the full conditionals $\pi(\boldsymbol{\varsigma}|Y^{(n)})$ and $\pi(w_t|Y^{(n)},\boldsymbol{\varsigma})$, $t = 1,\ldots,n$, and numerical integration routines to integrate out the hyperparameters $\boldsymbol{\varsigma}$.

In order to do so, INLA first approximates $\pi(\boldsymbol{\varsigma}|Y^{(n)})$, using a Gaussian approximation to the full conditional distribution of $\boldsymbol{w} = \{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_t\}$, by a multivariate Gaussian density $\tilde{\pi}_G(\boldsymbol{w}|Y^{(n)},\boldsymbol{\varsigma})$ evaluated at its mode $\boldsymbol{w}^{\star}(\boldsymbol{\varsigma})$. Then the posterior density of $\boldsymbol{\varsigma}$ is approximated by using the Laplace approximation (Tierney and Kadane (1986))

$$\tilde{\pi}\big(\boldsymbol{\varsigma}|Y^{(n)}\big) \propto \left.\frac{\pi(\boldsymbol{w},Y^{(n)},\boldsymbol{\varsigma})}{\tilde{\pi}_G(\boldsymbol{w}|Y^{(n)},\boldsymbol{\varsigma})}\right|_{\boldsymbol{w}=\boldsymbol{w}^{\star}(\boldsymbol{\varsigma})}. \tag{17}$$

The second step is to compute the Laplace approximation of $\pi(w_{tj}|Y^{(n)},\boldsymbol{\varsigma})$ for selected values of $\boldsymbol{\varsigma}$, which will be used to perform a numerical integration to obtain the posterior marginals of w_{tj} presented in (15). The density $\pi(w_{tj}|Y^{(n)},\boldsymbol{\varsigma})$ is approximated by

$$\tilde{\pi}_{LA}\big(w_{tj}|Y^{(n)},\boldsymbol{\varsigma}\big) \propto \left.\frac{\pi(\boldsymbol{w},Y^{(n)},\boldsymbol{\varsigma})}{\tilde{\pi}_G(\boldsymbol{w}_{-tj}|w_{tj},Y^{(n)},\boldsymbol{\varsigma})}\right|_{\boldsymbol{w}_{-tj}=\boldsymbol{w}^{\star}_{-tj}(w_{tj},\boldsymbol{\varsigma})}, \tag{18}$$

where $\boldsymbol{w}_{-tj}$ denotes the vector $\boldsymbol{w}$ without the jth component of vector $\boldsymbol{w}_t$, $\tilde{\pi}_G(\boldsymbol{w}_{-tj}|w_{tj},Y^{(n)},\boldsymbol{\varsigma})$ is the Gaussian approximation of $\pi(\boldsymbol{w}_{-tj}|w_{tj},Y^{(n)},\boldsymbol{\varsigma})$, treating w_{tj} as observed and $\boldsymbol{w}^{\star}_{-tj}(w_{tj},\boldsymbol{\varsigma})$ is the mode of $\pi(\boldsymbol{w}_{-tj}|w_{tj},Y^{(n)},\boldsymbol{\varsigma})$.

The approximation $\tilde{\pi}_{LA}(w_{tj}|Y^{(n)},\boldsymbol{\varsigma})$ in (18) can be quite expensive, since it is necessary to recompute $\tilde{\pi}_G(\boldsymbol{w}_{-tj}|w_{tj},Y^{(n)},\boldsymbol{\varsigma})$ for all w_{tj} and $\boldsymbol{\varsigma}$. Rue, Martino and Chopin (2009) propose two alternatives to obtain these full conditionals in a cheaper way. We focus our analysis in the simplified Laplace approximation defined as the third order Taylor expansion of $\tilde{\pi}_{LA}(w_{tj}|Y^{(n)},\boldsymbol{\varsigma})$ around w_{tj} and approximated by a skew-normal distribution (for more details see Rue, Martino and Chopin (2009)).

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

$$\begin{aligned}\pi\big(w_{tj}|Y^{(n)}\big) &= \int \pi\big(w_{tj}|Y^{(n)},\boldsymbol{\varsigma}\big)\pi\big(\boldsymbol{\varsigma}|Y^{(n)}\big)\,d\boldsymbol{\varsigma}\\ &\approx \sum_l \tilde{\pi}\big(w_{tj}|Y^{(n)},\varsigma_l\big)\tilde{\pi}\big(\varsigma_l|Y^{(n)},\big)\Delta_l,\end{aligned}$$

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, Martino and Chopin (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.

The INLA procedure directly calculates the posterior marginal of interest making posterior inference over $\boldsymbol{\psi}$ very efficient. More details can be found in Rue, Martino and Chopin (2009).

3.2 Inference in the observation driven model

In the observation driven approach, the only quantity to be estimated is the parameter vector $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\phi}', \boldsymbol{\theta}')$. In the GARMA model, the log-likelihood function is conditional on the first $s > \max(p, q)$ observations and on $\ln(y_t^*)$ for $t = 1, \ldots, \max(p, q)$. For the GLARMA model, the likelihood function is computed by conditioning on values of $u_0, u_{-1}, \ldots, u_{1-q}$ being equal to zero.

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

$$l_P\big(\boldsymbol{\delta}; Y^{(n)}\big) \propto \sum_{t=s+1}^{n} \big\{\big(\boldsymbol{x}_t'\boldsymbol{\beta} + Z_t\big)y_t - e^{\boldsymbol{x}_t'\boldsymbol{\beta}+Z_t}\big\}, \tag{19}$$

while for the negative binomial it is given by

$$l_{NB}\big(\boldsymbol{\delta}; Y^{(n)}\big) \propto \left\{ \sum_{t=s+1}^{n} \big(\big(\boldsymbol{x}_t'\boldsymbol{\beta} + Z_t\big)y_t - \ln\big(r + e^{(\boldsymbol{x}_t'\boldsymbol{\beta}+Z_t)}\big)\big) \right\}, \tag{20}$$

where Z_t is given by (9) for GLARMA and by (14) for GARMA.

For GLARMA models, Davis, Dunsmuir and Streett (2003) provide parameter estimation through maximum likelihood, using the Newton–Raphson algorithm. In Davis, Dunsmuir and Streett (2005), the authors show the asymptotic normality of the maximum likelihood estimator (MLE) of $\boldsymbol{\delta}$ when $\lambda = 1$. For the GARMA approach, Benjamin, Rigby and Stasinopoulos (2003) compute the MLE for $\boldsymbol{\delta}$ using iteratively reweighted least squares and partial likelihood. Maximization of $l(\boldsymbol{\delta}; Y^{(n)})$ is done using the Fisher scoring algorithm.

In this work, the estimation is performed using MCMC algorithms. The Bayesian approach initially requires the specification of a prior distribution, $\pi(\boldsymbol{\delta})$, for the parameter vector $\boldsymbol{\delta}$. One possibility for $\pi(\boldsymbol{\delta})$ is to use normal priors with zero mean and variance σ_ϕ^2, σ_θ^2 and σ_β^2 for $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, respectively. Assuming independence among the parameters, the joint prior is

$$\pi(\boldsymbol{\delta}) \propto \frac{1}{(\sigma_\phi)^p(\sigma_\theta)^q(\sigma_\beta)^m} \exp\left[-\frac{\sum_{i=1}^{p}\phi_i^2}{2\sigma_\phi^2} - \frac{\sum_{i=1}^{q}\theta_j^2}{2\sigma_\theta^2} - \frac{\sum_{k=1}^{m}\beta_k^2}{2\sigma_\beta^2} \right]. \tag{21}$$

A simpler alternative, obtained as the limit of the prior distribution in (21), is the uniform prior, given by $\pi(\boldsymbol{\delta}) = c \in \Re$ for all possible values of $\boldsymbol{\delta}$, and 0 otherwise.

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

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

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

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

As the posterior distribution does not have a closed form, a Metropolis–Hastings (M-H) algorithm will be employed to estimate the parameter vector. The chosen kernel is based on a random walk chain, with a multivariate normal density,

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

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 $\boldsymbol{\delta}$ 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 (1996), 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 (MC) 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 (R Core Team (2018)). 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).

Series of sizes $n = 100$, 500 and 1000 were generated under each one of the models described in Section 2, and the three models were fitted to all series. As the results do not have a significant change for different sample sizes, only the $n = 100$ case is presented here. 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 for 1000 MC replications, considering two structures for the mean process, μ_t: a pure autoregressive or an autoregressive plus a covariate. In all cases, the parameter estimators are the posterior means.

4.1 Model with a pure autoregressive in the mean

Series were generated for the 3 scenarios with $\mu_t = \exp(Z_t)$, where Z_t in each model is given by

ZEG: $Z_t = \phi Z_{t-1} + u_t, u_t \overset{\text{i.i.d.}}{\sim} N(0,1)$.
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 $\lambda = 0.5$.

Table 1 shows the results for series of size $n = 100$. As expected, the estimates closer to the real value and the smallest mean squared error (MSE) 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 general, there is not one approach that is better in all situations. The ZEG and GARMA procedures seem to be more reliable when $\phi = 0.7$ and 0.95, and the GLARMA gives good estimates when $\phi = 0.4$ (except for the ZEG DGP). In fact, the GLARMA model undergo many problems for large values of ϕ. For example, when $\phi = 0.7$, the GLARMA approach seems to perform very well, but a close inspection revealed that the log-likelihoods for the generated series under this process can present local maximums far from the real value and only a very

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

		Real value of parameter ϕ					
		$\phi = 0.4$		$\phi = 0.7$		$\phi = 0.95$	
DGP	Fitted model	$\hat{\phi}$	MSE	$\hat{\phi}$	MSE	$\hat{\phi}$	MSE
ZEG	ZEG	**0.398**	**0.029**	**0.698**	**0.013**	**0.916**	**0.022**
	GARMA	0.195	0.073	**0.699**	0.045	0.973	0.103
	GLARMA	0.128	0.074	0.152	0.295	0.043	0.754
GARMA	ZEG	0.001	0.157	0.783	0.211	0.915	0.139
	GARMA	**0.397**	**0.008**	**0.693**	**0.009**	**0.933**	**0.008**
	GLARMA	0.409	0.012	0.638	0.018	0.641	0.137
GLARMA	ZEG	0.001	0.144	0.840	0.025	**0.952**	**0.012**
	GARMA	0.017	0.135	0.836	**0.023**	0.984	0.028
	GLARMA	**0.398**	**0.006**	**0.683**	0.043	0.049	0.696

Obs.: In bold are the estimates closer to the real values and the smallest MSE.

few points around the exact value. Thus, it is very difficult to reach the global maximum, unless the initial value in the optimization algorithm is very close to the real value.

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. 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 (once again, results not presented, but available upon request), residual analysis does not show, in general, any inadequacies concerning the assumptions of zero mean and homoscedasticity. 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 procedure always produces autocorrelated residuals. 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.

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

Series of size $n = 100$ were generated 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: $Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + u_t, u_t \overset{\text{i.i.d.}}{\sim} N(0, 1)$.
GARMA: $Z_t = \sum_{i=1}^{p} \phi_i (\ln(y^*_{t-i}) - \alpha - \beta x_{t-i})$, 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 mean squared error of the conditional mean (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 1 presents boxplots for the estimation of α, β and ϕ under the three DGP's. The first conclusion we can draw 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 for the ZEG DGP, 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, Kukuk and Liesenfeld (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.

Estimates of ϕ are shown at the bottom of Figure 1. 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 $\phi = 0.4$ 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 2 shows mean estimates of MSE($\hat{\mu}$) and PMSE for forecasts $h = 1$, 5 and 15 steps-ahead for each procedure. 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.

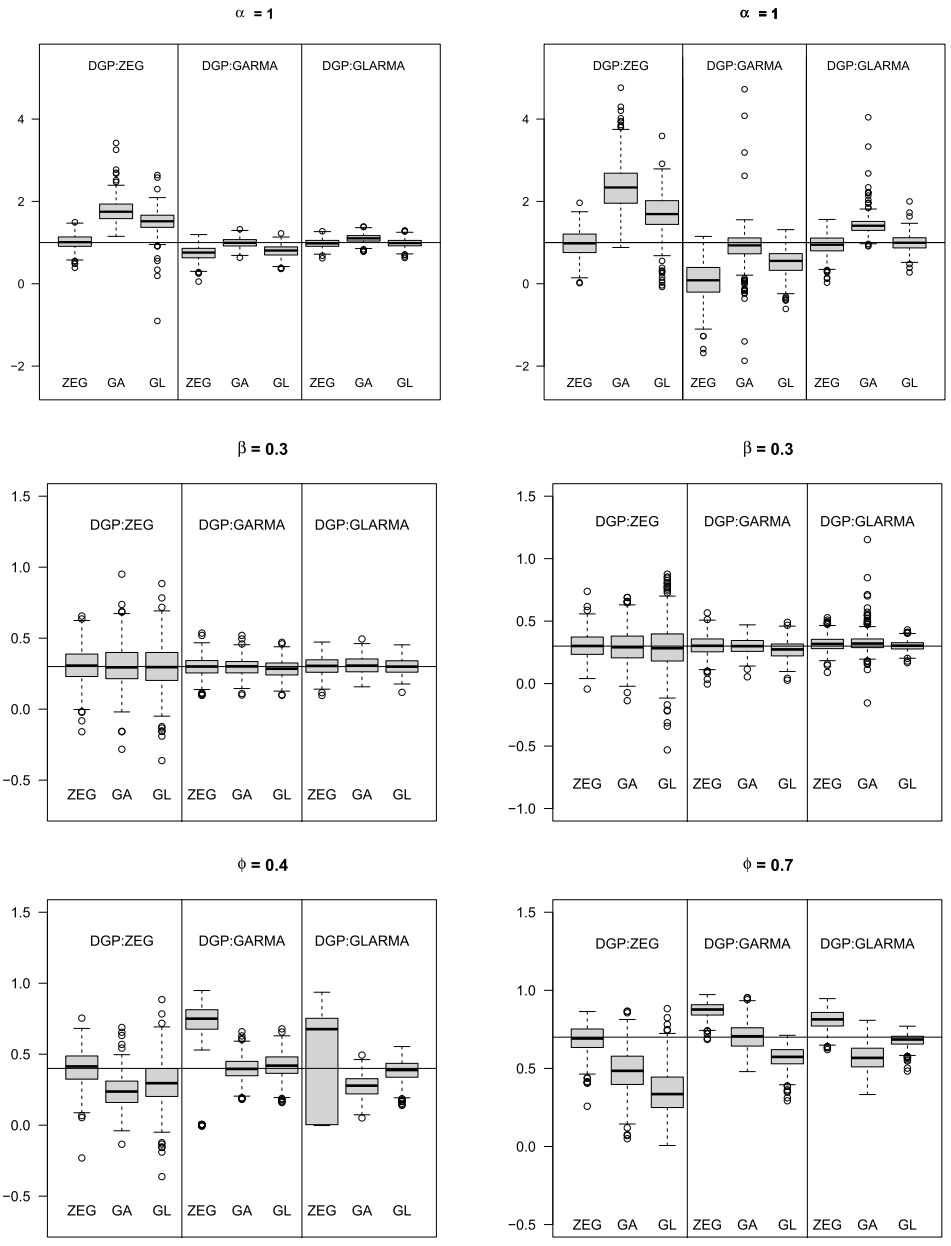

Figure 1 *Boxplot for the estimates of* α, β *and* ϕ_1, *in* 1000 *MC replications, for series of size* $n = 100$. *GA means GARMA and GL means GLARMA. A straight horizontal line is placed on the real value of the parameters. In the left-hand side are the results for models generated with* $\phi = 0.4$, *while the right-hand side presents the results for* $\phi = 0.7$.

Table 2 *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**
		$\phi = 0.7$			
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.

Table 3 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 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 when $\phi_1 = 0.50$, $\phi_2 = 0.25$. 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.

5 Application to the bankruptcy series

This example refers to the UCLA-LoPucki Bankruptcy Research Database, a set of monthly data of bankruptcy in United States companies, in the period from

Table 3 *Mean and MSE 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	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)
	GARMA	0.294	**0.295**	0.158	0.147	1.705	0.298	0.346	0.185
		(1.034)	(0.019)	(0.019)	(0.021)	(0.772)	(0.024)	(0.044)	(0.115)
	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)
GARMA	ZEG	0.353	0.302	1.420	−0.536	0.353	**0.302**	0.956	−0.158
		(1.061)	(0.011)	(097)	(0.347)	(0.149)	(0.010)	(0.573)	(0.275)
	GARMA	**0.966**	**0.301**	**0.253**	**0.236**	**0.819**	0.307	**0.477**	0.229
		(0.030)	(0.005)	(0.006)	(0.006)	(0.396)	(0.007)	(0.010)	(0.082)
	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)
GLARMA	ZEG	0.962	0.311	1.420	−0.536	0.980	**0.303**	0.935	−0.280
		(0.090)	(0.003)	(0.886)	(0.652)	(0.017)	(0.004)	(0.822)	(0.419)
	GARMA	1.148	0.306	0.169	0.160	1.411	0.314	0.406	0.143
		(0.037)	(0.003)	(0.012)	(0.014)	(0.207)	(0.003)	(0.027)	(0.140)
	GLARMA	**0.982**	**0.300**	**0.240**	**0.234**	**1.010**	**0.303**	**0.464**	**0.238**
		(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.

Jan/1980 to Sep/2014. This dataset contains bankruptcy cases on all American public companies that have declared assets of more than US$100 million (measured in 1980 dollars), the year before the firm went bankrupt. The covariates comprise macroeconomic and financial data obtained online from the Federal Reserve Bank of St. Louis (https://www.stlouisfed.org/).

The Poisson and negative binomial observation and parameter driven models were used to fit the number of companies that went bankrupt, shown in Figure 2(a). The last 12 observations were excluded from the fit with the purpose of comparing the forecasts. The final set of covariates, after excluding non-significant variables, were the Moody's Seasoned Baa Corporate Bond Yield (BAA) and 10-Year Treasury Constant Maturity Rate (GS10).

Figure 2(a) also reveals some peaks around 1991, 2001 and 2009. The last peak can be credited to the Financial crisis of 2007–2010, which has caused a large number of companies going bankrupt, such as Lehman Brothers (Sep/2008), Chrysler (Apr/2009) and General Motors (Jun/2009), among others. The increased number of bankruptcy in 2001 can be explained by the Western energy crisis of 2000–2001.

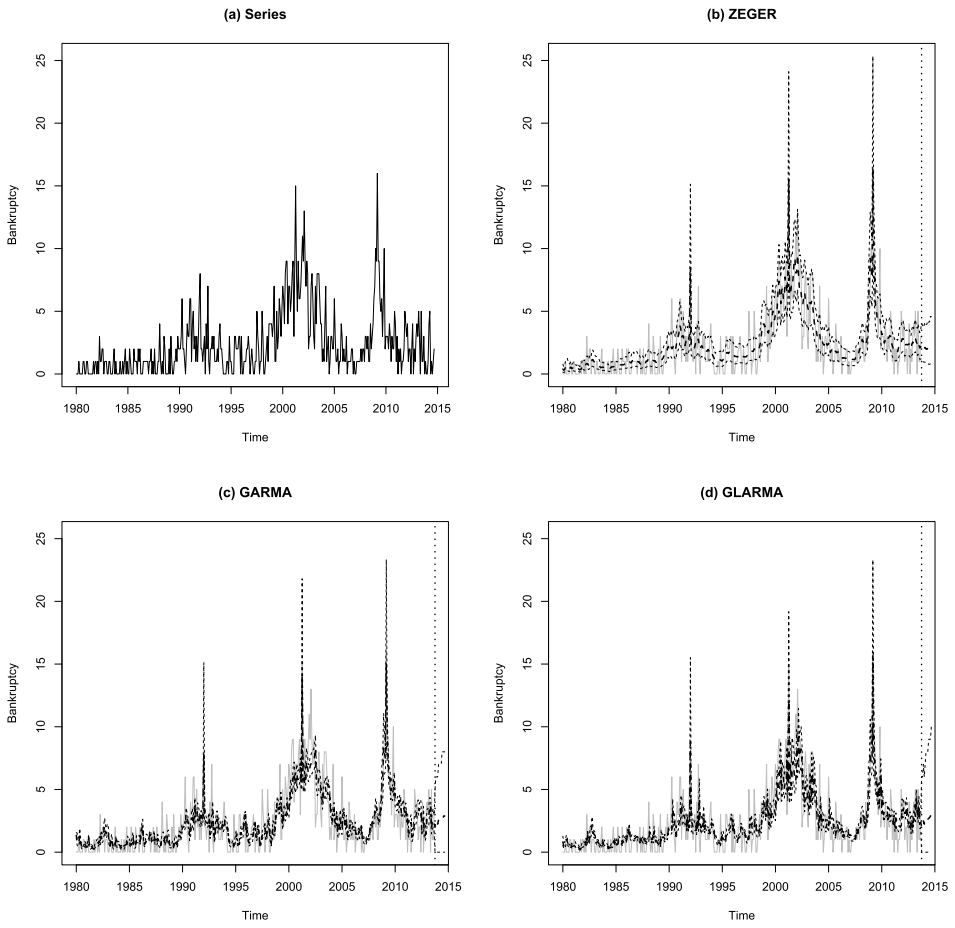

Figure 2 *Fitted models and forecasts for the Bankruptcy series. Full and dashed lines represent, respectively, the Bankruptcy time series and the fitted model. Dotted lines represent* 95% *credibility intervals. The vertical line separates the fitted data points from the forecast horizon.*

This crisis has affected mainly the state of California, and has produced large-scaled blackouts and an increase of electricity prices, causing the collapse of some important energy companies, such as the Pacific Gas & Electric Co (Apr/2001) and Enron (Dec/2001). With respect to the large number of bankrupt around 1991, a possible explanation is that, in this period, credit was very tight so it was difficult for firms to refinance debt. Thus, it was not the recession, but the overspending that has caused the main bankrupt cases at that time. These events were added to the model using dummy variables (Ind1991, Ind2001 and Ind2009) to account for interventions in these periods.

Table 4 shows the fit of the Poisson and negative binomial models to the data. These two models, under the three approaches (ZEG, GARMA and GLARMA), are compared using the DIC statistic. We can perceive that the Poisson model presented the smallest values of DIC for all approaches. Comparing the observation and parameter driven models, the ZEG procedure presented the smallest DIC, either for the Poisson or the negative binomial.

The fit was very similar for both Poisson and negative binomial models. The estimates of the parameters related to the covariates BAA and GS10 were significant for all models. The coefficient for BAA, an indicator of the interest rate, showed a positive relation with the number of companies that went bankrupt, while GS10, used as a reference for pricing debt securities issued by entities such as corporations and institutions, presented a negative relationship. The dummy variable Ind1991 was significant for all models, the Ind2001 was significant in the GARMA and negative binomial GLARMA models and the Ind2009 was only significant in the GLARMA model. An AR process of order 6 was included in order to whiten the residuals, although some lags were not significant for the ZEG and GARMA models. The precision parameter, r, was estimated with large values for the negative binomial regression, which corroborates with the similar estimates obtained in both regression analyses. As it is known, a large r estimate indicates that there is no need for overdispersion when modeling the data and thus the negative binomial model recovers the Poisson.

The residual analysis for the two models are very similar and, since the estimate of r was large and the Poisson presented the smallest DICs, we only show the results for this model (see Figure 3). The assumptions of independence and constant variance seem to be achieved in the observation models, while the ACF and PACF plots of the ZEG model show some peaks in the first lags. This behaviour of the ZEG model was already pointed out in Section 4.1 and Zeger (1988).

The fitted models and forecasts for the last 12 observations (Oct/2013 to Sep/2014) are presented again only for the Poisson model (Figures 2(b), 2(c) and 2(d)). It can be seen that the ZEG procedure produces estimates for the process mean which seem more smoothed, compared to the observation driven models. Regarding the predictions, the GARMA and GLARMA models present an increasing trend for the future observations, while the ZEG procedure shows a decreasing behaviour, with smaller credibility intervals.

The prediction mean square error (PMSE) was also computed (see Table 4). The negative binomial presented smaller PMSE than the Poisson model only for the ZEG approach, even though the values are fairly comparable. However the opposite occur in the GARMA and GLARMA cases, where the Poisson model presented smaller PMSE. Regarding the three procedures, the ZEG presented the smallest PMSE for both Poisson and negative binomial.

Table 4 *Poisson and negative binomial fit for the bankruptcy series*

	Poisson			Negative Binomial		
Model	ZEG	GARMA	GLARMA	ZEG	GARMA	GLARMA
Intercept	−0.18 (−1.91; 1.17)	**1.10** (0.37; 1.85)	0.87 (0.11; 1.70)	−0.01 (−1.29; 0.23)	**1.17** (0.42; 1.96)	0.80 (−0.14; 1.66)
BAA	**0.39** (0.20; 0.57)	**0.29** (0.13; 0.43)	**0.29** (0.09; 0.46)	**0.40** (0.22; 0.57)	**0.28** (0.11; 0.44)	**0.30** (0.10; 0.51)
GS10	**−0.44** (−0.60; −0.28)	**−0.35** (−0.48; −0.22)	**−0.43** (−0.60; −0.25)	**−0.47** (−0.63; −0.31)	**−0.35** (−0.50; −0.20)	**−0.44** (−0.63; −0.26)
Ind1991	**1.11** (0.23; 1.90)	**0.89** (0.10; 1.54)	**0.95** (0.28; 1.56)	**1.11** (0.02; 2.18)	**0.94** (0.14; 1.65)	**1.00** (0.16; 1.78)
Ind2001	0.62 (−0.03; 1.23)	**0.88** (0.33; 1.34)	0.58 (−0.02; 1.11)	0.73 (−0.20; 1.69)	**0.93** (0.35; 1.52)	**0.60** (0.04; 1.27)
Ind2009	0.53 (−0.11; 1.13)	0.49 (−0.06; 0.96)	**0.56** (0.14; 0.99)	0.54 (−0.42; 1.52)	0.52 (−0.12; 1.12)	**0.61** (0.04; 1.15)
AR(1)	**1.26** (0.02; 0.98)	**0.22** (0.14; 0.29)	**0.20** (0.14; 0.27)	**0.48** (0.99; 1.52)	**0.22** (0.16; 0.29)	**0.20** (0.14; 0.26)
AR(2)	**−0.49** (−1.13; −0.09)	**0.16** (0.08; 0.23)	**0.10** (0.04; 0.15)	0.23 (−0.41; 0.50)	**0.16** (0.08; 0.23)	**0.10** (0.03; 0.16)
AR(3)	**0.55** (0.15; 1.41)	0.06 (−0.01; 0.12)	**0.11** (0.05; 0.18)	0.17 (−0.24; 0.41)	0.05 (−0.02; 0.12)	**0.11** (0.04; 0.18)

Table 4 (*Continued*)

	Poisson			Negative Binomial		
Model	ZEG	GARMA	GLARMA	ZEG	GARMA	GLARMA
AR(4)	−0.17 (−0.92; 0.31)	0.03 (−0.04; 0.10)	**0.07** (0.01; 0.12)	0.18 (−0.36; 0.33)	0.03 (−0.04; 0.10)	**0.06** (0.01; 0.12)
AR(5)	−0.66 (−1.20; 0.29)	**0.11** (0.03; 0.18)	**0.06** (0.01; 0.12)	0.17 (−0.22; 0.43)	**0.11** (0.04; 0.19)	**0.07** (0.00; 0.13)
AR(6)	0.50 (−0.20; 0.85)	**0.08** (0.01; 0.15)	**0.09** (0.03; 0.15)	0.14 (−0.10; 0.43)	**0.08** (0.01; 0.15)	**0.09** (0.03; 0.15)
r				**10.46** (6.59; 15.76)	**26.24** (24.51; 27.80)	**19.05** (14.20; 25.41)
DIC	1330.56	1338.42	1384.64	1375.07	1388.32	1384.97
PMSE	3.328	4.115	3.902	3.289	4.258	4.715

Obs.: In bold are the significant parameters.

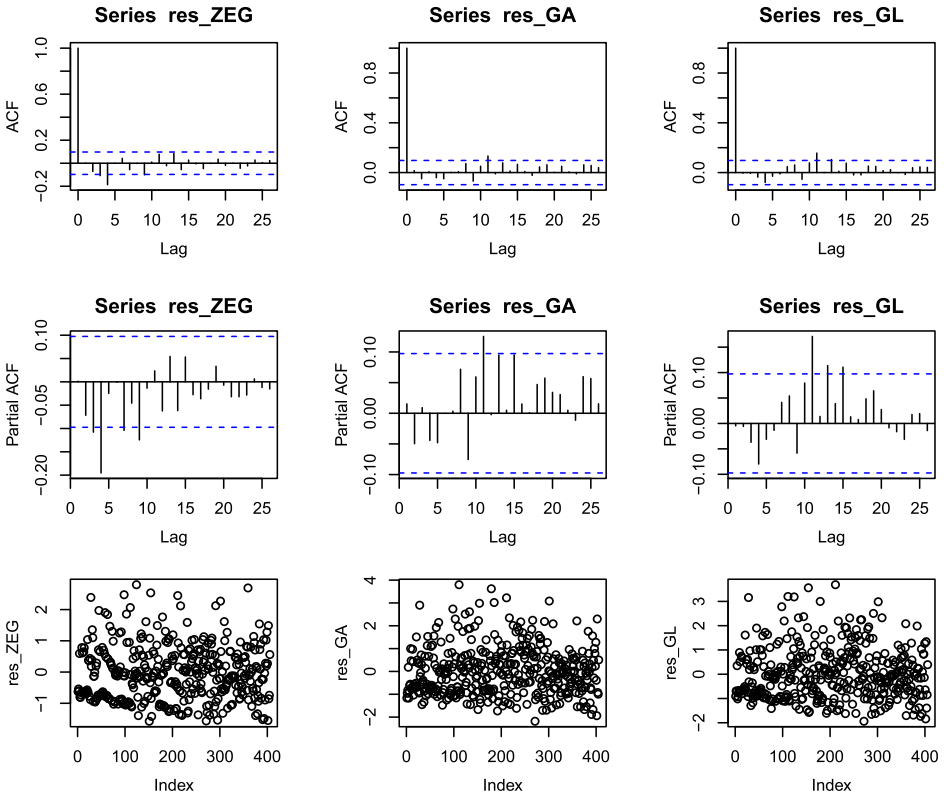

Figure 3 *Residual analysis for the three models fitted to the Bankruptcy series* (*Poisson*).

6 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. We believe this is the first attempt to compare the two approaches based on a Monte Carlo experiment.

Moreover, a Bayesian approach based on MCMC to estimate the parameters of the observation driven models, GARMA (Benjamin, Rigby and Stasinopoulos (2003)) and GLARMA (Davis, Dunsmuir and Streett (2003)) and an INLA approach to estimate the parameters of a model with a latent autorregressive process in the mean (Zeger (1988)) were introduced in the paper.

An extensive simulation study was performed in models with and without a covariate, using the Poisson distribution. The main conclusion that can be drawn is that the parameter of the covariate is not affected by the model chosen to fit the data, regardless the process which generates the data. This can be an indication that the methods are robust with respect to the estimation of this parameter. 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.

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 present an autocorrelated structure.

Acknowledgments

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 and CNPq/Brazil for partial financial support.

References

Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **72**, 1–33. MR2758115

Benjamin, M. A., Rigby, R. A. and Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association* **98**, 214–223. MR1965687

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day. MR0436499

Chan, K. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* **90**, 242–251. MR1325132

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics* **8**, 93–115. MR0623586

Creal, D., Koopman, S. J. and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* **28**, 777–795. MR3083971

Creal, D. D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews* **31**, 245–296.

Czado, C. and Kolbe, A. (2004). Empirical study of intraday option price changes using extended count regression models. Germany. Collaborative Research Center 386, Discussion Paper 403. Ludwig-Maximilians-Universitat.

Davis, R. A., Dunsmuir, W. T. M. and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika* **90**, 777–790. MR2024757

Davis, R. A., Dunsmuir, W. T. M. and Streett, S. B. (2005). Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodology and Computing in Applied Probability* **7**, 149–159. MR2189786

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

Diebold, F. X. and Schuermann, T. (1996). Exact maximum likelihood estimation of observation-driven econometric models. NBER Technical Working Paper No. 194.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* **89**, 603–615. MR1929166

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall. MR2260716

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**, 625–645. MR3127211

Gelman, A. (1996). Inference and monitoring convergence. In *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds.) 131–143. London: Chapman and Hall. MR1397966

Goudie, R. J. B., Turner, R. M., De Angelis, D., Thomas, A. and MultiBUGS (2017). A parallel implementation of the BUGS modelling framework for faster Bayesian inference. Available at https://arxiv.org/abs/1704.03216.

Harvey, A. C. and Fernandez, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics* **7**, 407–417.

Jung, R. C., Kukuk, M. and Liesenfeld, R. (2006). Time series of count data: Modeling, estimation and diagnostics. *Computational Statistics & Data Analysis* **51**, 2350–2364. MR2307505

Jung, R. C. and Tremayne, A. R. (2011). Useful models for time series of counts or simply wrong ones?. *AStA Advances in Statistical Analysis* **95**, 59–91. MR2775582

Koopman, S. J., Lucas, A. and Scharth, M. (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics* **98**, 97–110.

Li, W. K. (1994). Time series models based on generalized linear models: Some further results. *Biometrics* **50**, 506–511.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.

Meyer, R. and Yu, J. (2000). BUGS for a Bayesian analisys of stochastic volatility models. *Econometrics Journal* **3**, 198–215.

Migon, H. S., Gamerman, D., Lopes, H. F. and Ferreira, M. A. R. (2005). *Dynamic Models*. *Handbook of Statistics* **25**. Amsterdam: Elsevier. MR2490539

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

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, Statistical Methodology* **71**, 319–392. MR2649602

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

Shephard, N. and Pitt, M. K. (1995). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653–667. MR1603940

Smith, R. L. and Miller, J. E. (1986). A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **48**, 79–88. MR0848053

Souza, M. A. O. and Migon, H. S. (2018). Extended dynamic generalized linear models: The two-parameter exponential family. *Computational Statistics & Data Analysis* **121**, 164–179. MR3759205

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86. MR0830567

West, M., Harrison, J. and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association* **80**, 73–83. MR0786598

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika* **75**, 621–629. MR0995107

Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics* **44**, 1019–1031. MR0980997

G. C. Franco
M. O. Prates
Department of Statistics
Universidade Federal de Minas Gerais
Belo Horizonte, MG, 31270-901
Brazil
E-mail: glauraf@ufmg.br

H. S. Migon
Department of Statistical Methods
Universidade Federal do Rio de Janeiro
RJ
Brazil