

The Negative Binomial Process: A Tractable Model with Composite Likelihood-Based Inference

Wagner Barreto-Souza^{*‡*} and Hernando Ombao^{*†}

^{*}*Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

[‡]*Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil*

March 15, 2021

Abstract

We propose a log-linear Poisson regression model driven by a stationary latent gamma autoregression. This process has negative binomial (NB) marginals to analyze overdispersed count time series data. Estimation and statistical inference are performed using a composite (CL) likelihood function. We establish theoretical properties of the proposed count model, in particular, the strong consistency and asymptotic normality of the maximum CL estimator. A procedure for calculating the standard error of the parameter estimator and confidence intervals is derived based on the parametric bootstrap. Monte Carlo experiments were conducted to study and compare the finite-sample properties of the proposed estimators. The simulations demonstrate that, compared to the approach that combines generalized linear models with the ordinary least squares method, the proposed composite likelihood approach provides satisfactory results for estimating the parameters related to the correlation structure of the process, even under model misspecification. An empirical illustration of the NB process is presented for the monthly number of viral hepatitis cases in Goiânia (capital and largest city of the Brazilian state of Goiás) from January 2001 to December 2018.

Keywords: Bootstrap; Composite likelihood estimation; Count time series; Gamma autoregression; Negative binomial distribution; Overdispersion; Poisson regression.

^{*}Corresponding Author. E-mail: wagner.barretosouza@kaust.edu.sa

[†]E-mail: hernando.ombao@kaust.edu.sa

1 Introduction

Count data with serial correlation is ubiquitous. For example, epidemiologists monitor monthly cases of diseases (Zeger, 1988; Davis *et al.*, 2000), traffic and safety experts record accident counts (Brännäs & Johansson, 1994; Quddus, 2008) and sociologists track monthly counts of registered crime (Ristić *et al.*, 2009; Barreto-Souza, 2015).

To analyze count time series data, a popular class of models is the INteger-valued AutoRegres-sive (INAR) or more generally INARMA (INteger-valued AutoRegressive Moving Average), based on the binomial thinning operator (Steutel & van Harn, 1979), which were proposed and studied by McKenzie (1985, 1986), Al-Osh & Alzaid (1987), Alzaid & Al-Osh (1988), Al-Osh & Aly (1992), Freeland & McCabe (2004), and Enciso-Mora *et al.* (2009a). For a review on thinning-based count models, see Scotto *et al.* (2015). Variants of this model can be obtained, for instance, by assuming different operators rather than the binomial one. In this direction, see Aly & Bouzar (1994), Ristić *et al.* (2009), and Zhu & Joe (2010), just to name a few.

Another way for dealing with count time series data is to consider the INteger-valued Gener-alized AutoRegressive Conditional Heterokedastic (INGARCH) models by Ferland *et al.* (2006), Fokianos *et al.* (2009), Fokianos (2011), Zhu (2011), Fokianos & Fried (2012), Christou & Fokianos (2015), Davis & Liu (2016), and Silva & Barreto-Souza (2019), which are integer-valued counter-parts of the classical GARCH models by Bollerslev (1986).

There are other approaches to analyzing count time series data such as the Discrete ARMA (DARMA) model by Jacobs & Lewis (1983), the Generalized Linear Autoregressive Moving Average (GLARMA) process by Davis *et al.* (2003), and the renewal process-based count model by Cui & Lund (2009). A sample of references includes Davis *et al.* (1999), Kedem & Fokianos (2002), and Weiß (2018). We also recommend the very recent survey paper by Weiß (2020), where an updated account on stationary count time series models is provided.

In this paper, we develop a new approach inspired by the seminal work of Zeger (1988) that proposed a log-linear Poisson regression model conditioned on a stationary latent autoregressive process, where inference is performed through a quasi-likelihood method. This log-linear Poisson process belongs to the class of parameter-driven models (Cox, 1981). Many papers aimed to explore the inferential aspects of the log-linear Poisson model by Zeger (1988) such as Brännäs & Johansson (1994) and Davis *et al.* (2000). In these works and many others in the literature, the latent process is commonly assumed to follow a first-order Gaussian autoregressive process. Under

this assumption, the marginals of the counting process are Poisson log-normal distributed. The advantages of the estimation methods proposed by these authors are that the parameters related to the mean are estimated efficiently and closed forms for the asymptotic standard errors are derived. On the other hand, a limitation of these approaches is that it holds only for estimating efficiently the parameters related to the mean but not for latent process. This is a non-trivial limitation because the role of the latent process is to characterize the correlation structure which is crucial for prediction and forecasting.

Davis & Wu (2009) proposed a negative binomial regression with a latent process similarly as in Zeger (1988) by replacing the Poisson assumption with a negative binomial distribution. The model parameters were estimated by combining a generalized linear model (GLM) approach with ordinary least squares type of estimators. By assuming a strongly mixing latent process with a finite second moment, the GLM estimator for the mean parameters is shown to be asymptotically normally distributed. The latent process was assumed to follow a Gaussian AR(1) model in the numerical illustrations and extensions for the one-parameter exponential family were discussed. This extension unifies the Poisson log-normal and negative binomial logit regression models and opens the possibility to obtain new models.

Although the models based on the Zeger's approach can be easily defined, one major obstacle is that an explicit form for the likelihood function cannot be easily obtained. A solution for this was proposed by Chan & Ledolter (1995) who argued that, due to the very complex likelihood functions of these parameter-driven models, even an EM-algorithm is complicated to obtain because the E-step involves high-dimensional integration. With this, they proposed a Monte Carlo EM-algorithm for estimating the parameters. A drawback of this method is its high computational cost. Furthermore, the conditions for the Monte Carlo EM sequence of estimators to converge to the maximum likelihood estimator are not easily verifiable.

In this paper, we introduce a log-linear Poisson regression model driven by a stationary latent gamma autoregression with composite likelihood-based inference. The marginals of our proposed count model are negative binomial distributed, which we shall call the negative binomial process (BNP). The mixing between Poisson and gamma distributions in the iid (independent and identically distributed) setting is well-known and has been used for a long time to obtain a closed-form probability function, in this case, the negative binomial distribution. Here, we will demonstrate that this can be extended to the dynamic setting, which is one of the novelties of this paper. Below we list the main contributions of the proposed count time series model.

- (a) A remarkable feature of the negative binomial model is its tractability. We derive an explicit expression for its pair-joint probability functions, which enables us to make inference based on a composite likelihood function. With this, we can efficiently estimate the parameters of the latent process, which overcomes an existing problem in the approaches by [Zeger \(1988\)](#), [Davis *et al.* \(2000\)](#), and [Davis & Wu \(2009\)](#), where these are considered as nuisance parameters. We illustrate this advantage in this paper through a Monte Carlo simulation.
- (b) We provide the conditional distribution of our process at time $t + j$ given the process at time t , for $t, j \in \mathbb{N}$, which is not possible for instance in [Zeger \(1988\)](#), [Davis *et al.* \(2000\)](#) and [Davis & Wu \(2009\)](#). This is an advantage of the NB model because this conditional distribution is of great interest in time series analysis, for example, to perform prediction and propose diagnostic tools. Furthermore, we obtain analytic expressions for the conditional mean and variance of our process. Our approach can be adapted to obtain similar expressions for the models given in the three papers above mentioned. The basic properties of the latent process needed here (first two conditional moments) are well-known for the log-Gaussian autoregressive process (for instance, see [Davis & Wu \(2009\)](#)).
- (c) Under some conditions, we establish the consistency and asymptotic normality of the composite likelihood estimators of the negative binomial time series of counts.
- (d) The maximum composite likelihood (MCL) method proposed for our NB process has a low computational cost in contrast with the Monte Carlo EM-algorithm, MCMC, importance sampling, among others. Further, it delivers efficient model inference, which is demonstrated theoretically and empirically via simulations and real data application. Indeed a major advantage of the proposed estimation method is that it does not require full likelihood specification since it is based on pairwise probability functions.
- (e) We illustrate through a Monte Carlo simulation that the NB process can provide satisfactory results even under model misspecification, due to the MCL approach. This simulated study also shows the superior performance of the MCL estimation method over the GLM approach combined with ordinary least squares proposed by [Davis *et al.* \(2000\)](#) and [Davis & Wu \(2009\)](#).

Related work was developed in [Christou & Fokianos \(2014\)](#) where a dynamic mean was considered through a negative binomial INGARCH model. Estimation is carried out using a quasi-likelihood approach for the mean parameters and a type of method of moments for the dispersion.

In our case, the dispersion is estimated jointly with the other parameters through a composite likelihood approach. In other words, we do not assume the dispersion as a nuisance parameter since its efficient estimation may be of great interest in real data applications.

It is worth to note that the proposed NB process can be considered as a count stochastic volatility model and hence is a natural alternative to the class of INGARCH models. This relationship will be made more precise later. These approaches will be considered in the empirical illustration in Section 5. The inference on the INGARCH processes considered here is based on the quasi-likelihood method, combined with the method of moments, as proposed by Christou & Fokianos (2014) and implemented in the R package `tscount` by Liboschik *et al.* (2017). The composite likelihood inference considered in our paper is an alternative inferential procedure to the well-used quasi-likelihood method in the literature.

The remainder of this paper is organized in the following manner. In Section 2 the negative binomial process (NBP) is defined and its properties developed. Composite likelihood-based inference for our NB process is addressed in Section 3. Further, we discuss how to obtain standard errors of the parameter estimates and confidence intervals based on a parametric bootstrap approach. We also provide prediction under the NB process and propose a diagnostic tool based on the probability integral transform histogram for checking model adequacy. Two alternative methods of estimation additional to the CL approach are also presented. In Section 4 we present some Monte Carlo simulations to compare the finite-sample behavior of the proposed estimators and to study our model under misspecification. An empirical illustration of the monthly number of viral hepatitis cases in Goiânia (capital and largest city of the Brazilian state of Goiás) from January 2001 to December 2018 is presented in Section 5. Concluding remarks and future research are discussed in Section 6. Proofs of the results given in this paper are presented in Appendix A1. This paper has a Supplementary Material containing some additional simulated results and an R code for fitting the proposed model. An example illustrating the usage of our R codes is provided in Appendix A2.

2 Model specification

We begin this section with some preliminaries and notations. Here, denote a random variable Z following a gamma distribution with moment generating function $\Psi_Z(t) = E(\exp\{tZ\}) = (1 - t/\gamma)^{-\nu}$, for $t < \gamma$, to be $Z \sim G(\nu, \gamma)$, where $\nu > 0$ and $\gamma > 0$ are the shape and rate parameters,

respectively. The density function of Z assumes the form $f(z) = \frac{\gamma^\nu}{\Gamma(\nu)} z^{\nu-1} e^{-\gamma z}$, for $z > 0$. A random variable N following a Poisson distribution with mean $\lambda > 0$ is denoted by $N \sim \text{Pois}(\lambda)$. To define our proposed count process with negative binomial marginals, we present the gamma autoregressive (GAR) process introduced in [Sim \(1990\)](#). Here, the mean of the GAR process is fixed to 1 to avoid the non-identifiability problem when defining our negative binomial model in the sequence; please see [Remark 2.2](#) for more details.

Definition 2.1. (*Gamma AutoRegression*) *The first-order gamma autoregression (GAR) (with marginal mean 1) is defined by a Markov process, say $\{Z_t\}_{t \in \mathbb{N}}$, with transition probability density function assuming the form*

$$f(z_t|z_{t-1}) = \frac{\phi}{1-\rho} \left(\frac{z_t}{\rho z_{t-1}} \right)^{\frac{\phi-1}{2}} \exp\left(-\phi \frac{z_t + \rho z_{t-1}}{1-\rho}\right) I_{\phi-1}\left(2\phi \frac{\sqrt{\rho z_t z_{t-1}}}{1-\rho}\right), \quad z_t > 0,$$

for $z_{t-1} > 0$, $\phi > 0$, $\rho \in (0, 1)$, and $t \geq 1$, where $I_\nu(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{2k+\nu}}{\Gamma(k+\nu+1)k!}$ is the modified Bessel function of the first kind of order $\nu \in \mathbb{R}$, for $x \in \mathbb{R}$, and $Z_0 \sim G(\phi, \phi)$.

The GAR process depends on the parameters ϕ and ρ responsible for controlling the dispersion and dependence, respectively. By taking $\rho \rightarrow 0^+$, we obtain that $\{Z_t\}_{t \in \mathbb{N}}$ is a sequence of iid random variables following $G(\phi, \phi)$ distribution.

The autoregressive process in [Definition 2.1](#) is stationary; the marginal is gamma with mean 1 and variance ϕ^{-1} . Denote $\{Z_t\} \sim \text{GAR}(1)$ which implicitly depends on the parameters ϕ and ρ . The joint density function of (Z_{t+j}, Z_t) is given by

$$f(z_{t+j}, z_t) = \frac{\phi^{\phi+1}}{(1-\rho^j)\Gamma(\phi)} \left(\frac{z_{t+j} z_t}{\rho^j} \right)^{\frac{\phi-1}{2}} \exp\left(-\phi \frac{z_{t+j} + z_t}{1-\rho^j}\right) I_{\phi-1}\left(2\phi \frac{\sqrt{\rho^j z_{t+j} z_t}}{1-\rho^j}\right), \quad (1)$$

for $z_{t+j}, z_t > 0$ and $t, j \in \mathbb{N}$, where $I_\nu(\cdot)$ is the modified Bessel function defined above.

Further, the correlation of the GAR(1) process is $\text{corr}(Z_{t+j}, Z_t) = \rho^j$ and the conditional mean and variance are

$$E(Z_{t+j}|Z_t = z) = 1 + \rho^j(z - 1) \quad \text{and} \quad \text{Var}(Z_{t+j}|Z_t = z) = \frac{1-\rho^j}{\phi} [1 + \rho^j(z - 1)], \quad (2)$$

for $z > 0$. The above results on the GAR(1) process are due to [Sim \(1990\)](#). In the following proposition, we state an important property of the GAR process which was not discussed in the

paper by [Sim \(1990\)](#).

Proposition 2.1. *The gamma autoregressive process in (2.1) is strongly mixing.*

Proof. Since the GAR(1) process is strictly stationary and Markovian (for instance, see [Sim \(1990\)](#)), by using Corollary 2.2 from [Chernick \(1981\)](#), it is enough to show that the joint density function of (Z_{t+j}, Z_t) converges pointwise to the product of the marginal densities of Z_{t+j} and Z_t , as $j \rightarrow \infty$, for all $t \in \mathbb{N}$. We have that $I_{\phi-1} \left(2\phi \frac{\sqrt{\rho^j z_1 z_2}}{1 - \rho^j} \right) = \frac{\phi^{\phi-1}}{\Gamma(\phi)} \left(\frac{\rho^j z_1 z_2}{1 - \rho^j} \right)^{\frac{\phi-1}{2}} + O(\rho^{j(2+\phi)})$, as $j \rightarrow \infty$, where $I_{\phi-1}(\cdot)$ is the modified Bessel function of the first kind of order $\phi - 1$ defined after Eq. (1).

Hence, for $z_1, z_2 > 0$, it follows that

$$\begin{aligned} \lim_{j \rightarrow \infty} f_{Z_{t+j}, Z_t}(z_1, z_2) &= \lim_{j \rightarrow \infty} \left\{ \frac{\phi^{2\phi}}{\Gamma(\phi)^2 (1 - \rho^j)^{\frac{\phi+1}{2}}} (z_1 z_2)^{\phi-1} \exp \left(-\phi \frac{z_1 + z_2}{1 - \rho^j} \right) + O(\rho^{j(2+\phi)}) \right\} \\ &= \frac{\phi^\phi}{\Gamma(\phi)} z_1^{\phi-1} \exp(-\phi z_1) \times \frac{\phi^\phi}{\Gamma(\phi)} z_2^{\phi-1} \exp(-\phi z_2), \end{aligned}$$

that is the product of two densities of gamma random variables with mean 1 and variance ϕ^{-1} . Therefore, we obtain the desired result. \square

The above property of the GAR process will be important in this paper to establish the consistency and asymptotic normality of the estimators based on the composite likelihood function in Section 3.

With the above ingredients, we now define our proposed negative binomial process (NBP) as follows.

Definition 2.2. *(Negative Binomial Process) Let $\{Z_t\}_{t \in \mathbb{N}}$ be a GAR(1) process as given in Definition 2.1 with parameters $\phi > 0$ and $\rho \in (0, 1)$. We define the negative binomial process, say $\{Y_t\}_{t \in \mathbb{N}}$, by assuming that $\{Y_t\}_{t \in \mathbb{N}}$ are conditionally independent given the latent gamma process $\{Z_t\}_{t \in \mathbb{N}}$ and that the following stochastic representation is satisfied: $Y_t | \{Z_j\}_{j \in \mathbb{N}} \sim \text{Pois}(Z_t \mu_t)$, for $t \in \mathbb{N}$, where $\mu_t = \exp(\mathbf{x}_t^\top \boldsymbol{\beta})$, \mathbf{x}_t is a p -dimensional covariate vector and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$ is the associated vector of regression coefficients. We denote $\{Y_t\} \sim \text{NBP}$ which implicitly depends on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi, \rho)^\top$.*

Remark 2.2. *We here discuss some points regarding the definition of our NB model given above. The assumption $E(Z_t) = 1 \forall t$ is necessary in order to avoid non-identifiability problems. A similar assumption is required for the existent models under the Zeger's approach. For instance, see [Davis](#)*

et al. (2000) and Davis & Wu (2009). We choose to work with the mean-based parametrization of the negative binomial since we can directly model the mean of the process through covariates, which provides a natural interpretation of the regression coefficients for practitioners.

Remark 2.3. *Our model can be considered as a count stochastic volatility process since $\text{Var}(Y_t|Z_t) = Z_t\mu_t$ (conditional variance is driven by a latent process); see Kim et al. (1998) for a reference on a continuous stochastic volatility model. Therefore, our proposed negative binomial process is an alternative to the INGARCH processes, since the continuous stochastic volatility and GARCH models are competitor approaches (e.g., Broto & Ruiz (2004)). More generally speaking, the models based on Zeger's approach are count stochastic volatility models. The NBP and the Poisson and NB INGARCH models will be compared in the hepatitis data analysis.*

In the following proposition, some basic quantities of the NB model are developed. In particular, the marginals of this process are negative binomial distributed.

Proposition 2.4. *Let $\{Y_t\}_{t \in \mathbb{N}}$ be a NB process. Then its marginal probability function, mean, variance and autocorrelation function can be derived, respectively, as follows:*

$$\Pr(Y_t = y) = \frac{\Gamma(y + \phi)}{y! \Gamma(\phi)} \left(\frac{\mu_t}{\mu_t + \phi} \right)^y \left(\frac{\phi}{\mu_t + \phi} \right)^\phi, \quad y, t \in \mathbb{N},$$

$$E(Y_t) = \mu_t, \quad \text{Var}(Y_t) = \mu_t(1 + \phi^{-1}\mu_t), \quad t \in \mathbb{N}, \quad \text{and}$$

$$\text{corr}(Y_{t+j}, Y_t) = \frac{\rho^j}{\sqrt{(1 + \phi/\mu_{t+j})(1 + \phi/\mu_t)}}, \quad t, j \in \mathbb{N}.$$

Proof. The expressions for the marginal probability function and the first two cumulants follow directly from the model definition. The expression for the correlation is obtained by following the same steps as Davis et al. (2000), and therefore its proof is omitted. \square

Remark 2.5. *We call the attention that our model has the ability to capture only the positive correlation since $\rho > 0$. A possible extension of our model to include negative autocorrelation is discussed in Section 6. Further, we have that Z_t converges in probability to 1 as $\phi \rightarrow \infty$ for all t .*

Consequently, $\lim_{\phi \rightarrow \infty} \text{corr}(Y_{t+j}, Y_t) = \lim_{\phi \rightarrow \infty} \frac{\rho^j}{\sqrt{(1 + \phi/\mu_{t+j})(1 + \phi/\mu_t)}} = 0$, regardless of the value of

ρ . This is a more general feature of models based on Zeger's approach where decreasing variance in the latent process implies decreasing correlation in the observation process.

Some joint and conditional properties of the negative binomial time series are now developed. These will be important later when discussing inferential aspects. We present the joint probability function of the bivariate vector (Y_{t+j}, Y_t) in the following proposition (the proof is given in the Appendix).

Proposition 2.6. For $t, j \in \mathbb{N}$, we have that the joint probability function of the bivariate vector (Y_{t+j}, Y_t) , say $p(y, x) = Pr(Y_{t+j} = y, Y_t = x)$, is given by

$$p(y, x) = \frac{\phi^{2\phi}(1 - \rho^j)^{\phi+x+y} \mu_t^x \mu_{t+j}^y \Gamma(x + \phi) \Gamma(y + \phi)}{\Gamma(\phi)^2 x! y! [(1 - \rho^j)\mu_t + \phi]^{x+\phi} [(1 - \rho^j)\mu_{t+j} + \phi]^{y+\phi}} {}_2F_1(x + \phi, y + \phi; \phi; \lambda_{tj}),$$

for $x, y \in \mathbb{N}$, where ${}_2F_1(a, b; c, d)$ is the hypergeometric function defined for $a, b, c > 0$ and $|d| < 1$ and $\lambda_{tj} = \frac{\phi^2 \rho^j}{[(1 - \rho^j)\mu_t + \phi][(1 - \rho^j)\mu_{t+j} + \phi]}$.

The next proposition gives the first two conditional cumulants of Y_{t+j} given Y_t . The proof of this proposition can be found in the Appendix.

Proposition 2.7. The conditional mean and variance of Y_{t+j} given Y_t are respectively

$$E(Y_{t+j}|Y_t) = \mu_{t+j} \left(1 + \rho^j \frac{Y_t - \mu_t}{\mu_t + \phi} \right) = \mu_{t+j} \frac{\mu_t(1 - \rho^j) + \phi + \rho^j Y_t}{\mu_t + \phi} \quad \text{and}$$

$$Var(Y_{t+j}|Y_t) = \mu_{t+j}(1 - \rho^j) \left(1 + \mu_{t+j} \frac{1 - \rho^j}{\phi} \right) + \mu_{t+j} \rho^j \left(1 + 2\mu_{t+j} \frac{1 - \rho^j}{\phi} + \frac{\mu_{t+j}}{\mu_t + \phi} \rho^j \right) \frac{Y_t + \mu_t}{\mu_t + \phi},$$

for $t, j \in \mathbb{N}$.

In what follows, we discuss parameter estimation under the NB process based on a composite likelihood approach and establish the consistency and asymptotic normality of the proposed estimators.

3 Composite likelihood inference

Let y_1, \dots, y_n be observed values of a negative binomial time series of counts Y_1, \dots, Y_n according Definition 2.2. We now discuss inference for our NB model based on the composite likelihood (CL) approach. For a review on the CL inference, refer to Varin (2008) and Varin et al. (2011). Let

$\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \rho)^\top$ be the parameter vector. The inference here will be based on the following composite likelihood, which is also known in the literature as the pairwise likelihood (PL) of order $m \in \mathbb{N}$:

$$\text{PL}^{(m)}(\boldsymbol{\theta}) = \prod_{t=m+1}^n \prod_{j=1}^m p(y_t, y_{t-j}),$$

where $p(y, x) = \Pr(Y_t = y, Y_{t-j} = x)$ is the joint probability function given in Proposition 2.6, for $x, y \in \mathbb{N}$. The associated log-likelihood function is denoted by $\ell_n^{(m)}(\boldsymbol{\theta}) = \log \text{PL}^{(m)}(\boldsymbol{\theta})$.

The maximum composite likelihood (MCL) estimator of $\boldsymbol{\theta}$ is obtained as the global maximizer of the log-composite likelihood function, that is $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \ell_n^{(m)}(\boldsymbol{\theta})$. There is no closed form for the estimator and in this paper we adopted the Nelder-Mead optimization method (Nelder & Mead, 1965). All the numerical results presented in this paper were obtained by using the R program (R Development Core Team, 2016).

Let $\mathbf{U}(\boldsymbol{\theta}) = \nabla \ell_n^{(m)}(\boldsymbol{\theta})$ be the score function associated to the CL function $\ell_n^{(m)}(\boldsymbol{\theta})$. The standard errors of the parameter estimates can be obtained through the Godambe information (Godambe, 1960) denoted here by $\mathbf{G}(\boldsymbol{\theta})$:

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})^{-1}\mathbf{H}(\boldsymbol{\theta}), \tag{3}$$

where $\mathbf{H}(\boldsymbol{\theta}) = E(-\nabla \mathbf{U}(\boldsymbol{\theta}))$ and $\mathbf{J}(\boldsymbol{\theta}) = \text{var}(\mathbf{U}(\boldsymbol{\theta}))$. The composite likelihood methods are usually less efficient compared to full likelihood methods (for instance, see Vidoni (2018)). On the other hand, one advantage is that they might work better (yielding consistent estimates) under misspecification since full likelihood specification is not required.

We now establish the strong consistency and asymptotic normality of the maximum composite likelihood estimators for the NB model. For this, we first need to state some assumptions on the covariates.

Assumptions 1 (AS1): *The covariates are bounded and composed by a trend function $f(t/n)$, where $f(\cdot)$ is a continuous function on $[0, 1]$ (n is the sample size and $t = 1, \dots, n$), and also by harmonic functions that describe the seasonal effects.*

Assumptions 2 (AS2): *Here, we consider that the covariates are stationary, m -dependent, and bounded. More specifically, by $\{\mathbf{X}_t\}_{t \in \mathbb{N}}$ m -dependent we mean that there is an $m \in \mathbb{N}$ such that any finite-dimensional marginals $(\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_n})$ and $(\mathbf{X}_{j_1}, \dots, \mathbf{X}_{j_k})$ with $j_1 - i_n > m$ are independent, where $i_1 < \dots < i_n$ and $j_1 < \dots < j_k$ are sets of indexes.*

Remark 3.1. Assumptions AS1 are given in [Davis et al. \(2000\)](#) and [Davis & Wu \(2009\)](#) while Assumptions AS2 are considered in [Ng et al. \(2011\)](#). From these papers, under these conditions, we apply the limit theorems for strongly mixing processes to establish the desired asymptotic properties, as will be seen in the proof of the theorem. Other types of covariates rather than those given in AS1 and AS2 can be considered, but it is necessary to impose certain conditions to establish asymptotic properties of the MCL estimators.

Theorem 3.2. Let Θ be a compact parameter space containing the true parameter vector value, denoted by $\theta_0 = (\beta_0, \phi_0, \rho_0)^\top$, and assume that the covariates satisfy either (AS1) or (AS2). Then, we have that $\hat{\theta} \xrightarrow{a.s.} \theta_0$ and $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\mathbf{0}, \Sigma)$ as $n \rightarrow \infty$, where Σ is the asymptotic covariance matrix whose form is specified below.

Proof. We begin by proving the strong consistency of the MCL estimator $\hat{\theta}$. Define $\tilde{\ell}_n^{(m)}(\theta) = n^{-1}\ell^{(m)}(\theta)$ and $\tilde{\ell}^{(m)}(\theta) = E\left(\sum_{j=1}^m \log p(y_t, y_{t-j})\right)$. We now argue that $\tilde{\ell}^{(m)}(\theta)$ in fact exists for all $\theta \in \Theta$ and $\sup_{\theta \in \Theta} |\tilde{\ell}^{(m)}(\theta)| < K$, for some $K > 0$. Since $p(y_1, y_2) \leq 1$ for all $y_1, y_2 \in \mathbb{N}$, we have that $\tilde{\ell}^{(m)}(\theta) \leq 0$ for all $\theta \in \Theta$.

The expectation $\tilde{\ell}^{(m)}(\theta)$ is a linear combination (with positive coefficients) of expectation of terms like Y_t , $\log \Gamma(Y_t + \phi)$ and $\log {}_2F_1(Y_{t-j} + \phi, Y_t + \phi; \phi, \lambda_{tj})$. We have that $E_{\theta_0}(Y_t) = \exp(\mathbf{x}^\top \beta_0)$ is bounded (because the covariates are bounded) and $E_{\theta_0}(\log \Gamma(Y_t + \phi)) \geq E_{\theta_0}(\log Y_t!) > -\infty$. Further, since ${}_2F_1(Y_{t-j} + \phi, Y_t + \phi; \phi, \lambda_{tj}) > \inf_{\theta \in \Theta} \min_{x \in \mathbb{N}} \frac{\Gamma(\phi + x)^2}{\Gamma(\phi)} > 0$, we have that $E({}_2F_1(Y_{t-j} + \phi, Y_t + \phi; \phi, \lambda_{tj})) > -\infty$. Therefore, it follows that $\tilde{\ell}^{(m)}(\theta)$ exists.

By the Ergodic Theorem, we get $\tilde{\ell}_n^{(m)}(\theta) \xrightarrow{a.s.} \tilde{\ell}^{(m)}(\theta)$, as $n \rightarrow \infty$. With this result and by following the same steps of Lemma A.4 from [Ng et al. \(2011\)](#), we can establish the strong consistency $\hat{\theta} \xrightarrow{a.s.} \theta_0$.

Now, by expanding $\tilde{\ell}_n^{(m)}(\theta)$ in Taylor's Series around the true value θ_0 as discussed by [Pedeli & Varin \(2020\)](#), we obtain

$$\sqrt{n}(\hat{\theta} - \theta_0) = \left(-\nabla \tilde{U}(\theta_0)\right)^{-1} \sqrt{n}\tilde{U}(\theta_0) \{1 + O_p(1)\},$$

where we have defined $\tilde{U}(\theta) = \nabla \tilde{\ell}_n^{(m)}(\theta)$ and the above $O_p(1)$ term is due to the consistency of the MCL estimator discussed above.

Under either Condition AS1 or AS2 on the covariates, from Proposition 2.1 and Definition 2.2, we have that $\{Y_t, \dots, Y_{t-m}, Z_t, \dots, Z_{t-m}\}_{t \geq m+1}$ is a strong mixing Markov process (see Remark 3.1 and the related papers); the strong mixing property of $\{Y_t, \dots, Y_{t-m}\}$ follows from [Davis & Wu](#)

(2009) and Ng *et al.* (2011) under conditions AS1 and AS2, respectively. Hence, it follows from the Ergodic Theorem and the Central Limit Theorem that $-\nabla\tilde{U}(\boldsymbol{\theta}_0) \xrightarrow{a.s.} \boldsymbol{\Omega}_2$ and $\sqrt{n}\tilde{U}(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_1)$ as $n \rightarrow \infty$, respectively, for some matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, where $\boldsymbol{\Omega}_2$ is invertible. Therefore, we obtain $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$ as $n \rightarrow \infty$, with $\boldsymbol{\Sigma} = \boldsymbol{\Omega}_2^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}_2^{-1}$. \square

The computation of the Godambe information (3) is very cumbersome. This matrix can be estimated by using empirical estimators for the matrices $\mathbf{H}(\boldsymbol{\theta})$ and $\mathbf{J}(\boldsymbol{\theta})$ as discussed by Varin (2008). For our model, this approach is still cumbersome since the empirical estimators will depend on the gradient and hessian function of the log-likelihood $\ell^{(m)}(\boldsymbol{\theta})$, involving the first and second partial derivatives of the hypergeometric function, which may be computationally unstable. Due to the difficulty of computation, our approach is to use the parametric bootstrap procedure (Efron & Tibsharani, 1999) for computing the standard errors of the maximum composite likelihood estimates as follows.

Given the data, in the b -th bootstrap replicate, generate $Y_{1b}^*, \dots, Y_{nb}^*$ following a NB process with setting parameter vector $\hat{\boldsymbol{\theta}}$, for $b = 1, \dots, B$, with B being the total number of bootstrap. In each replicate, estimate the parameter vector by using the above maximum composite likelihood approach. Denote by $\hat{\boldsymbol{\theta}}_b^*$ the MCL estimate obtained in the b -th replicate of the parametric bootstrap, for $b = 1, \dots, B$. With this bootstrap sample, we obtain the standard errors of the parameter estimates and also construct confidence intervals.

We now discuss prediction and some diagnostic tools for the proposed NB model. From the results presented in Proposition 2.7, the k -step ahead prediction for the count mean at time $t + k$ given Y_t , say $\hat{Y}_{t+k,t}(\hat{\boldsymbol{\theta}}) \equiv \hat{E}(Y_{t+k}|Y_t)$, is given by

$$\hat{Y}_{t+k,t}(\hat{\boldsymbol{\theta}}) = \hat{\mu}_{t+k} \left(1 + \hat{\rho}^k \frac{Y_t - \hat{\mu}_t}{\hat{\mu}_t + \hat{\phi}} \right), \tag{4}$$

for $k \in \mathbb{N}$. Note that prediction of Y_{t+k} given Y_t requires that we know the covariate vector \mathbf{x}_{t+k} at time t . This is possible, for example, when the covariates are functions only on the time like trends and harmonic functions as those used in our simulations and application to a real data set.

We also call the attention that other predicted quantities can be obtained from the conditional distribution of Y_{t+k} given Y_t , which is fully specified with the results given in Propositions 2.6 and 2.7. Prediction intervals can be derived based on the parametric bootstrap described above. For some existing count time series models in the literature presented in Section 1, prediction is not discussed since an explicit expression for the conditional mean is not available. On the other hand,

an explicit expression for the conditional mean of such models can be obtained by following the steps of proof of Proposition 2.7.

Note that our process is not Markovian but rather a hidden Markov chain. Thus a question that naturally arises is whether or not one should use the entire history Y_1, \dots, Y_t rather than Y_t for forecasting or predicting Y_{t+k} , for $k \geq 1$. One way to take this into account under a composite likelihood perspective was recently discussed by Vidoni (2018). Assume that the parameters are known for simplicity of notation. In practice, we replace them by their consistent composite likelihood estimators. Following Vidoni (2018), we define the surrogate predictive composite likelihood function by

$$p_{pred}(y_{t+k}|y_t, \dots, y_1; \boldsymbol{\theta}) = c(y_t, \dots, y_1; \boldsymbol{\theta}) \prod_{i=0}^t p_{t+k-i}(y_{t+k}|y_i; \boldsymbol{\theta}), \tag{5}$$

where $c(y_t, \dots, y_1; \boldsymbol{\theta})$ is a normalized constant to get a valid predictive distribution, which is well-defined in our case since the terms in (5) are bounded, and

$$p(y_{t+k}|y_i; \boldsymbol{\theta}) = \frac{\phi^\phi (1 - \rho^{t+k-i})^{\phi+y_i+y_{t+k}} \mu_{t+k}^{y_{t+k}} \Gamma(y_{t+k} + \phi)}{\Gamma(\phi) y_{t+k}! [(1 - \rho^{t+k-i}) \mu_{t+k} + \phi]^{y_{t+k} + \phi}} \left(\frac{\mu_i + \phi}{(1 - \rho^{t+k-i}) \mu_i + \phi} \right)^{y_i + \phi} \times {}_2F_1(y_i + \phi, y_{t+k} + \phi; \phi, \lambda_{i \ t+k-i}), \tag{6}$$

with $\lambda_{i \ t+k-i}$ as in Proposition 2.6. In the empirical application, we will illustrate and compare the prediction methods discussed here. A particular advantage in using the method proposed by Vidoni (2018) is that it can be obtained as a byproduct of our composite likelihood inference approach. For readers interested in a great account on measures of forecast/prediction accuracy, we refer to Hyndman & Koehler (2006).

To check model adequacy, we consider the non-randomized Probability Integral Transform (PIT) histogram proposed by Czado *et al.* (2009), which has been widely used in count time series models. For simplicity of notation, we discuss the case $m = 1$ (order of the composite likelihood function) considered in both simulated and real data applications in this paper.

All our inference procedure is based on the pairwise probability functions. Therefore, model adequacy relies on the correct/suitable specification of these functions. Following Czado *et al.*

(2009), we define

$$F_t(u|y_t) = \begin{cases} 0, & \text{if } u \leq p_{t,t-1}(y_t - 1|y_{t-1}; \boldsymbol{\theta}), \\ \frac{u - p_{t,t-1}(y_t - 1|y_{t-1}; \boldsymbol{\theta})}{p_{t,t-1}(y_t|y_{t-1}; \boldsymbol{\theta}) - p_{t,t-1}(y_t - 1|y_{t-1}; \boldsymbol{\theta})}, & \text{if } p_{t,t-1}(y_t - 1|y_{t-1}; \boldsymbol{\theta}) \leq u < p_{t,t-1}(y_t|y_{t-1}; \boldsymbol{\theta}), \\ 1, & \text{if } u \geq p_{t,t-1}(y_t|y_{t-1}; \boldsymbol{\theta}), \end{cases}$$

for $0 \leq u \leq 1$, where $p_{t,t-1}(\cdot|\cdot; \boldsymbol{\theta})$ is defined in (6). Hence, the PIT is given by

$$\bar{F}(u) = \frac{1}{n} \sum_{t=2}^n F_t(u|y_t), \quad 0 \leq u \leq 1.$$

Note that $\bar{F}(\cdot)$ is the cumulative distribution function of a uniform distribution if the predictive distribution is correctly specified (Czado *et al.*, 2009). One method to verify if this uniformity is reasonable for the fitted model is to plot the histogram of the heights $\text{freq}_h = \bar{F}\left(\frac{h}{H}\right) - \bar{F}\left(\frac{h-1}{H}\right)$, for $h = 1, \dots, H$, where H is the number of bins. Typical choices for this number are $H = 10$ or $H = 20$. This procedure will be considered in the hepatitis data analysis for checking the model adequacy.

In the next subsection, we present additional alternative estimation methods. Results of the analysis from the different methods will be compared and contrasted through Monte Carlo simulation in Section 4.

3.1 Alternative estimation methods

We here discuss two additional methods of estimation for the NB time series model. More specifically, we consider a type-conditional least squares method and a generalized linear model approach, which have been used in the literature. As before, y_1, \dots, y_n denote observed time series of counts of Y_1, \dots, Y_n from the NB model.

Let us start with the conditional least squares (CLS) type of estimators, which are obtained by minimizing the Q -function given by

$$Q(\boldsymbol{\theta}) = \sum_{t=2}^n (y_t - E(Y_t|y_{t-1}))^2 = \sum_{t=2}^n \left[y_t - \mu_t \left(1 + \rho \frac{y_{t-1} - \mu_{t-1}}{\mu_{t-1} + \phi} \right) \right]^2.$$

A closed form expression for the estimators is hard to obtain and hence numerical optimization procedure is necessary. Computation is performed using the `optim` R function with the Nelder-

Mead method for this optimization problem. Here we give an explicit analytical form of the CLS estimator for ρ as a function of the other CLS estimators, that is

$$\hat{\rho}_{cls} = \frac{\sum_{t=2}^n \bar{\mu}_t (Y_t - \bar{\mu}_t) (Y_{t-1} - \bar{\mu}_{t-1}) / (\bar{\mu}_{t-1} + \hat{\phi}_{cls})}{\sum_{t=2}^n \bar{\mu}_t^2 (Y_{t-1} - \bar{\mu}_{t-1})^2 / (\bar{\mu}_{t-1} + \hat{\phi}_{cls})^2},$$

where $\bar{\mu}_t = \exp(\mathbf{x}_t^\top \hat{\boldsymbol{\beta}}_{cls})$, for $t = 1, \dots, n$, with $\hat{\boldsymbol{\beta}}_{cls}$ and $\hat{\phi}_{cls}$ being the CLS estimators of $\boldsymbol{\beta}$ and ϕ , respectively, which do not have closed forms.

The generalized linear model (GLM) estimator for the parameter vector $\boldsymbol{\beta}$ is obtained by maximizing the log-likelihood $\ell(\boldsymbol{\beta})$ given by

$$\ell(\boldsymbol{\beta}) = \sum_{t=1}^n \{y_t \log \mu_t - \mu_t - \log y_t!\},$$

which ignores the latent process; for more details, see [Davis et al. \(2000\)](#). We use the `glm` function from the R program to obtain the GLM estimates of $\boldsymbol{\beta}$. Assuming a linear Gaussian process and under some regularity conditions on the regressors, which holds for trend functions forming triangular arrays, harmonic functions and stationary processes, [Davis et al. \(2000\)](#) established the consistency and asymptotic normality of the GLM estimators. To estimate the parameters ρ and ϕ , we can proceed as [Davis & Wu \(2009\)](#) and use the ordinary least squares type of estimators suggested by [Brännäs & Johansson \(1994\)](#). In this way, we obtain the following estimators:

$$\hat{\phi}_{ols} = \frac{\sum_{t=1}^n \tilde{\mu}_t^4}{\sum_{t=1}^n \tilde{\mu}_t^2 \{(Y_t - \tilde{\mu}_t)^2 - \tilde{\mu}_t\}}$$

and

$$\hat{\rho}_{ols} = \hat{\phi}_{ols} \frac{\sum_{t=2}^n \tilde{\mu}_t \tilde{\mu}_{t-1} (Y_t - \tilde{\mu}_t) (Y_{t-1} - \tilde{\mu}_{t-1})}{\sum_{t=2}^n \tilde{\mu}_t \tilde{\mu}_{t-1}^2}, \tag{7}$$

where $\tilde{\mu}_t = \exp(\mathbf{x}_t^\top \hat{\boldsymbol{\beta}}_{glm})$, for $t = 1, \dots, n$, with $\hat{\boldsymbol{\beta}}_{glm}$ being the GLM estimator of $\boldsymbol{\beta}$.

Another approach to be considered using the negative binomial regression was suggested by one of the referees. From [Proposition 2.4](#), we have that the marginals of our process are negative

binomial distributed. Based on this, we can use the following log-likelihood function:

$$\tilde{\ell}(\boldsymbol{\beta}, \phi) \propto \phi \log \phi - \log \Gamma(\phi) + \sum_{t=1}^n \left\{ y_t \log \left(\frac{\mu_t}{\mu_t + \phi} \right) + \log \Gamma(y_t + \phi) - \phi \log(\mu_t + \phi) \right\}.$$

Therefore, estimators for $\boldsymbol{\beta}$ and ϕ can be obtained by maximizing the log-likelihood $\tilde{\ell}$. In fact, the `glm.nb` routine from the MASS library in the R program implements the estimation algorithm; denote these estimators by $\hat{\boldsymbol{\beta}}_{nb}$ and $\hat{\phi}_{nb}$. We propose an alternative estimator for ρ by using (7) with $\hat{\boldsymbol{\beta}}_{nb}$ and $\hat{\phi}_{nb}$ instead of $\hat{\boldsymbol{\beta}}_{glm}$ and $\hat{\phi}_{ols}$, respectively. This estimation method is shortly denoted by NB/OLS.

4 Monte Carlo simulation

In this section we evaluate the finite-sample behavior of the estimators proposed in the previous sections through Monte Carlo simulation. We consider the NB model under two configurations with moderate ($\rho = 0.5$) and high correlation ($\rho = 0.8$), and mean structure given by $\mu_t = \exp(\beta_0 + \beta_1 t/n + \beta_2 \cos(2\pi t/12))$, for $t = 1, \dots, n$, with sample sizes $n = 100, 200, 500, 1000$. More explicitly, we set $(\beta_0, \beta_1, \beta_2) = (-1, 0.5, 2)$ in both cases and consider $(\phi, \rho) = (1.5, 0.5)$ and $(\phi, \rho) = (0.5, 0.8)$ in the first and second configurations, respectively. The mean function contains both the linear temporal trend and seasonality component. These components intend to make the simulation more realistic because they mimic the actual features in many seasonal-related disease count data.

We ran Monte Carlo simulations with 1000 replications. Tables 1 and 2 present the empirical mean of the estimates with their associated root mean squared error (RMSE) based on the MCL, CLS, GLM/OLS and NB/OLS estimation methods for the first and second configurations, respectively. We here consider MCL estimation based on the pairwise likelihood function of order $m = 1$. Concerning the estimation of the parameter vector $\boldsymbol{\beta}$, we can observe from these tables that the MCL, GLM, and NB methods present similar results in terms of bias and RMSE, with a slight advantage for the MCL method. The poorest results are due to the CLS method. In particular, we observe that this method yields estimates with considerable bias.

For estimating ϕ , the MCL and NB methods present satisfactory and similar results, while the other methods do not perform well for small and moderate sample sizes. Upon inspection of the results regarding the estimation of ρ (which is also related to the latent process), we see

Table 1: Empirical mean of the estimates of the parameters and their associated RMSE based on the MCL, CLS, GLM/OLS, and NB/OLS estimation methods. We here set $(\beta_0, \beta_1, \beta_2) = (-1, 0.5, 2)$, $\phi = 1.5$ and $\rho = 0.5$.

		ESTIMATES				RMSE			
		MCL	CLS	GLM	NB	MCL	CLS	GLM	NB
$n = 100$	$\beta_0 = -1$	-1.068	-1.954	-1.078	-1.083	2.137	4.418	2.146	2.148
	$\beta_1 = 0.5$	0.523	0.528	0.528	0.548	1.398	2.869	1.402	1.386
	$\beta_2 = 2$	2.024	2.839	2.022	2.018	1.731	3.603	1.731	1.727
	$\phi = 1.5$	1.985	2.626	2.852	1.988	2.257	5.308	3.696	2.267
	$\rho = 0.5$	0.454	0.418	0.439	0.370	1.114	1.122	1.124	1.130
$n = 200$	$\beta_0 = -1$	-1.060	-1.437	-1.069	-1.068	2.086	2.835	2.097	2.093
	$\beta_1 = 0.5$	0.536	0.530	0.545	0.549	1.272	1.526	1.288	1.270
	$\beta_2 = 2$	2.018	2.396	2.020	2.016	1.700	2.408	1.704	1.699
	$\phi = 1.5$	1.752	2.140	2.236	1.762	1.727	4.277	2.547	1.742
	$\rho = 0.5$	0.463	0.429	0.426	0.381	1.088	1.118	1.112	1.114
$n = 500$	$\beta_0 = -1$	-1.005	-1.122	-1.007	-1.008	2.010	2.181	2.012	2.011
	$\beta_1 = 0.5$	0.487	0.470	0.488	0.491	1.138	1.205	1.148	1.138
	$\beta_2 = 2$	2.003	2.134	2.003	2.002	1.670	1.843	1.670	1.668
	$\phi = 1.5$	1.608	1.788	1.778	1.613	1.443	3.200	1.679	1.450
	$\rho = 0.5$	0.477	0.463	0.450	0.432	1.070	1.096	1.085	1.094
$n = 1000$	$\beta_0 = -1$	-1.006	-1.051	-1.007	-1.007	2.004	2.061	2.004	2.004
	$\beta_1 = 0.5$	0.503	0.496	0.504	0.504	1.083	1.122	1.089	1.084
	$\beta_2 = 2$	2.001	2.049	1.998	2.000	1.665	1.731	1.664	1.664
	$\phi = 1.5$	1.569	1.789	1.663	1.571	1.380	2.954	1.476	1.381
	$\rho = 0.5$	0.483	0.493	0.468	0.454	1.061	1.112	1.074	1.089

that the MCL method continues yielding satisfactory results in contrast with the other methods considered. Further, we observe that the NB/OLS method gives the worst performance, which yields considerable bias even for moderate sample sizes.

We also illustrate this fact through figures presented in the Supplementary Material, which give box-plots of the estimates of the parameters ϕ and ρ , respectively, based on the MCL, CLS, GLM/OLS and NB/OLS methods for the two configurations considered with $n = 100$ and $n = 500$. One possible explanation for why the CL approach produces lower bias and RMSE than the GLM/OLS (for ϕ and ρ) and NB/OLS (for ρ) alternatives is that it is a likelihood-based method while the others are based on a type-method of moments. It is well-known in the literature that likelihood approaches are more efficient compared to method of moments. Therefore, the superior performance of the composite likelihood approach over the other methods is already expected. We further observe from these figures a deviation from the normality in some cases.

In the Supplementary Material, we also provide additional simulated results to explore the asymptotic normality of the MCL estimators for finite-sample sizes. To summarize, we observe that the normal approximation for the MCL estimators for the β 's is quite satisfactory for all cases considered since there is a good agreement between the empirical distribution of the estimates and the normal distribution. Regarding the estimators of ϕ and ρ , some departures from the normal

Table 2: Empirical mean of the estimates of the parameters and their associated RMSE based on the MCL, CLS, GLM/OLS, and NB/OLS estimation methods. We here set $(\beta_0, \beta_1, \beta_2) = (-1, 0.5, 2)$, $\phi = 0.5$ and $\rho = 0.8$.

		ESTIMATES				RMSE			
		MCL	CLS	GLM	NB	MCL	CLS	GLM	NB
$n = 100$	$\beta_0 = -1$	-1.285	-3.495	-1.286	-1.302	2.406	7.523	2.397	2.419
	$\beta_1 = 0.5$	0.665	0.737	0.651	0.688	2.144	6.086	2.074	2.141
	$\beta_2 = 2$	2.023	3.666	2.026	2.020	1.835	4.933	1.842	1.833
	$\phi = 0.5$	0.791	0.735	1.511	0.809	1.138	2.511	2.269	1.158
	$\rho = 0.8$	0.690	0.500	0.601	0.402	1.008	1.023	1.022	1.015
$n = 200$	$\beta_0 = -1$	-1.179	-2.282	-1.180	-1.188	2.151	4.852	2.163	2.166
	$\beta_1 = 0.5$	0.581	0.604	0.572	0.592	1.784	4.048	1.807	1.807
	$\beta_2 = 2$	1.997	2.830	1.996	1.994	1.768	3.453	1.770	1.767
	$\phi = 0.5$	0.623	0.456	0.999	0.635	1.001	1.448	1.347	1.004
	$\rho = 0.8$	0.737	0.573	0.662	0.487	0.990	1.008	1.013	1.007
$n = 500$	$\beta_0 = -1$	-1.060	-1.333	-1.054	-1.064	1.959	2.486	1.960	1.968
	$\beta_1 = 0.5$	0.539	0.454	0.527	0.545	1.295	1.730	1.326	1.320
	$\beta_2 = 2$	2.001	2.257	1.996	1.999	1.741	2.157	1.737	1.740
	$\phi = 0.5$	0.544	0.436	0.690	0.550	0.962	1.129	0.998	0.962
	$\rho = 0.8$	0.782	0.686	0.751	0.621	0.986	0.993	0.990	0.985
$n = 1000$	$\beta_0 = -1$	-1.025	-1.123	-1.021	-1.026	1.880	2.062	1.881	1.883
	$\beta_1 = 0.5$	0.511	0.446	0.502	0.510	1.182	1.401	1.216	1.201
	$\beta_2 = 2$	1.994	2.098	1.992	1.993	1.732	1.883	1.730	1.731
	$\phi = 0.5$	0.522	0.487	0.605	0.526	0.965	1.090	0.983	0.965
	$\rho = 0.8$	0.790	0.739	0.771	0.687	0.986	0.993	0.988	0.984

distribution can be observed for the sample sizes in some cases, but for large sample sizes (for instance $n = 5000$), the normal approximation is suitable, which is in line with Theorem 3.2. For the cases where the asymptotic normality does not work well, we suggest the reparameterizations $\tilde{\phi} = \log \phi$ and $\tilde{\rho} = \text{logit}(\rho) = \log\left(\frac{\rho}{1-\rho}\right)$. In the Supplementary Material, we show that the asymptotic normality for the MCL estimators of the parameters $\tilde{\phi} = \log \phi$ and $\tilde{\rho}$ is quite satisfactory even for the sample size $n = 200$.

We now provide a small Monte Carlo simulation to check the performance of the bootstrap approach proposed for getting the standard errors of the MCL estimates. Since the asymptotic normality for the MCL estimators is adequate as discussed before, we consider the standard normal approach to construct the confidence intervals for the β 's. In Table 3 we provide the empirical coverages of the standard normal bootstrap confidence intervals for the β 's with confidence level $100(1 - \alpha)\%$, for $\alpha = 0.01, 0.05, 0.10$, and $n = 500$. We can observe a good agreement from the nominal and empirical coverages for all cases considered.

Regarding the confidence intervals for ϕ and ρ , we consider both quantile and standard normal bootstrap approaches. The simulated results for these cases are presented in Table 4, also for $n = 500$. From this table, we have that both approaches work well and similarly for $\rho = 0.5$ and

Table 3: Empirical coverages of the standard normal bootstrap confidence intervals for the β 's with confidence level $100(1 - \alpha)\%$, for $\alpha = 0.01, 0.05, 0.10$, and $n = 500$.

Conf. level	$(\phi, \rho) = (1.5, 0.5)$			$(\phi, \rho) = (0.5, 0.8)$		
	β_0	β_1	β_2	β_0	β_1	β_2
99%	0.990	0.987	0.982	0.970	0.983	0.989
95%	0.958	0.942	0.951	0.921	0.938	0.944
90%	0.902	0.887	0.900	0.882	0.878	0.898

Table 4: Empirical coverages of the quantile and standard normal bootstrap confidence intervals for ϕ and ρ with confidence level $100(1 - \alpha)\%$, for $\alpha = 0.01, 0.05, 0.10$, and $n = 500$.

Conf. level	Quantile				Standard normal			
	$(\phi, \rho) = (1.5, 0.5)$		$(\phi, \rho) = (0.5, 0.8)$		$(\phi, \rho) = (1.5, 0.5)$		$(\phi, \rho) = (0.5, 0.8)$	
	ϕ	ρ	ϕ	ρ	ϕ	ρ	ϕ	ρ
99%	0.981	0.988	0.963	0.971	0.996	0.987	0.993	0.997
95%	0.937	0.947	0.882	0.934	0.981	0.951	0.975	0.966
90%	0.880	0.901	0.797	0.895	0.960	0.913	0.937	0.931

$\rho = 0.8$. On the other hand, for $\phi = 1.5$ and $\phi = 0.5$ the best performances were achieved by the quantile and standard normal approaches, respectively.

Based on the previously simulated results, we make the following recommendations. Point estimation based on the MCL method yielded satisfactory results even for small sample sizes, which can be faced in practice. To obtain the standard errors and to construct confidence intervals, we recommend the bootstrap approach. More specifically, regarding the construction of confidence intervals, for the β 's and ρ unrestricted and for $\phi < 1$ we suggest the standard normal approach (being the quantile approach also adequate for ρ); the quantile approach is recommended for the parameter ϕ when $\phi \geq 1$.

We are now interested in investigating the performance of the NBP under misspecification. For this purpose, a small Monte Carlo simulation was conducted. We simulated Poisson autoregression with a latent log-normal AR(1) process having mean and variance (denoted here by σ^2) equal to 1 and correlation parameter (denoted by ρ) equal to 0.87. We consider a reparameterization of the Poisson log-normal (PLN) model such that the parameter ρ plays the same role that in our model. Similarly, we consider $1/\phi = \sigma^2$ in our model to be compared with the PLN process. The regression coefficients in both models also play the same role.

The mean structure considered was $\mu_t = \exp(\beta_0 + \beta_1 \cos(2\pi t/12))$, with $(\beta_0, \beta_1) = (1, 1)$ and $n = 100, 200, 500, 1000$. These settings were considered in [Davis et al. \(2000\)](#). We ran 1000 Monte Carlo replicas and considered the MCL (based on the NBP) and GLM/OLS estimation methods.

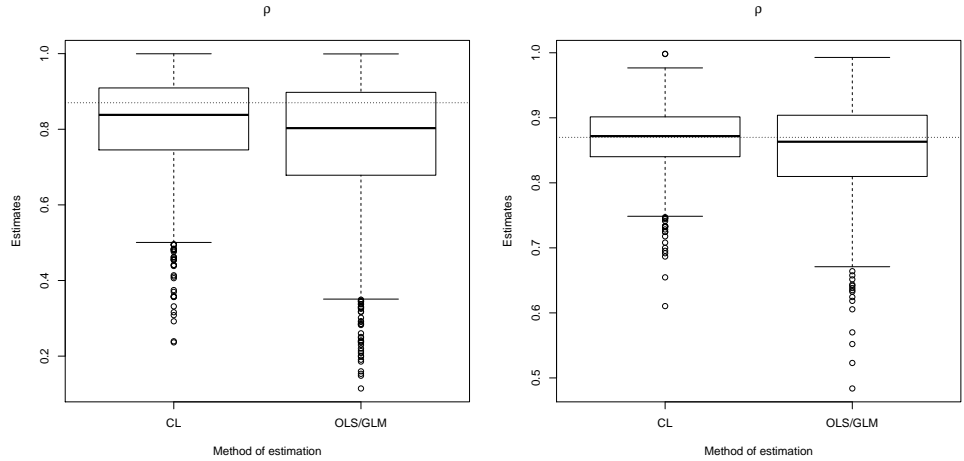


Figure 1: Box-plots of the estimates of the correlation parameter based on the MCL and GLM/OLS methods under the Poisson log-normal autoregression with $n = 100$ (on the left) and $n = 500$ (on the right); here $\sigma^2 = 1$ and $\rho = 0.87$.

The empirical means of the estimates of the parameters with their associated RMSEs are presented in the Supplementary Material to save space. Looking at these empirical means, we observe that both methods work very well for estimating the mean parameters. For the correlation parameter ρ , we see that the MCL method had superior performance over the GLM/OLS method. The NBP captures well the correlation even under the misspecification considered. This is illustrated in Figure 1, where boxplots of the estimates of the parameter ρ are presented. Looking at the results for the variance parameter, we have that the GLM/OLS method provides better results than the MCL method. With this small simulation, we see in general good performance of the NB model under misspecification.

5 Hepatitis data analysis

We now analyze the monthly number of viral hepatitis cases of residents in Goiânia (capital and largest city of the Brazilian state of Goiás) from January 2001 to December 2018, so yielding $n = 216$ observations. The data are available in the Supplementary Material. As far as we know, this very recent data has not been analyzed in the literature.

A plot of the number of hepatitis cases and the scatter plot of Y_t against Y_{t-1} for $t = 2, \dots, 216$ are presented in Figure 2. The time series plot (left) suggests a decreasing trend with some smooth

sinusoidal pattern in the mean count and the plot on the right indicates a strong linear dependence between the present and previous counts (empirical correlation equal to 0.702). The associated estimated autocorrelation (ACF) and partial autocorrelation (PACF) functions are displayed in Figure 3, which indicate autocorrelation and seasonality in the hepatitis data. Indeed one observes a high ACF estimate at lag 1 which is consistent with the observation in Figure 2.

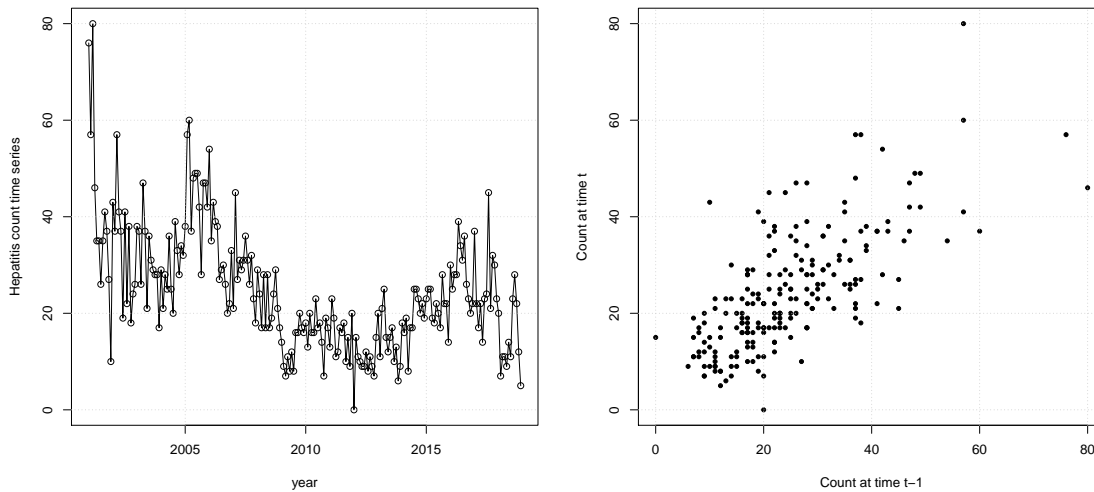


Figure 2: Plot of the monthly number of viral hepatitis cases of habitants in Goiânia from January 2001 to December 2018 (to the left) and scatter plot of Y_t against Y_{t-1} for $t = 2, \dots, 216$ (to the right).

We assume the following regression structure for the mean of our negative binomial process:

$$\mu_t = \exp \left\{ \beta_0 + \beta_1 t/216 + \beta_2 \cos(2\pi t/12) + \beta_3 \sin(2\pi t/12) + \beta_4 \cos(2\pi t/6) + \beta_5 \sin(2\pi t/6) + \beta_5 \cos(2\pi t/3) + \beta_6 \sin(2\pi t/3) \right\}, \quad t = 1, \dots, 216.$$

The components $\cos(2\pi t/12)$ and $\sin(2\pi t/12)$ capture the annual seasonality pattern in the log mean count while $\cos(2\pi t/6)$ and $\sin(2\pi t/6)$ deal with the semiannual seasonality. Quarter seasonality is captured from the components $\cos(2\pi t/3)$ and $\sin(2\pi t/3)$. The term $t/216$ is the linear trend, which is suggested in the plot in Figure 2.

In order to provide initial guesses for the maximum composite likelihood method, we consider the GLM/OLS approach. In Table 5, we present the maximum composite likelihood estimates of the parameters with their associated standard errors and confidence interval with a significance

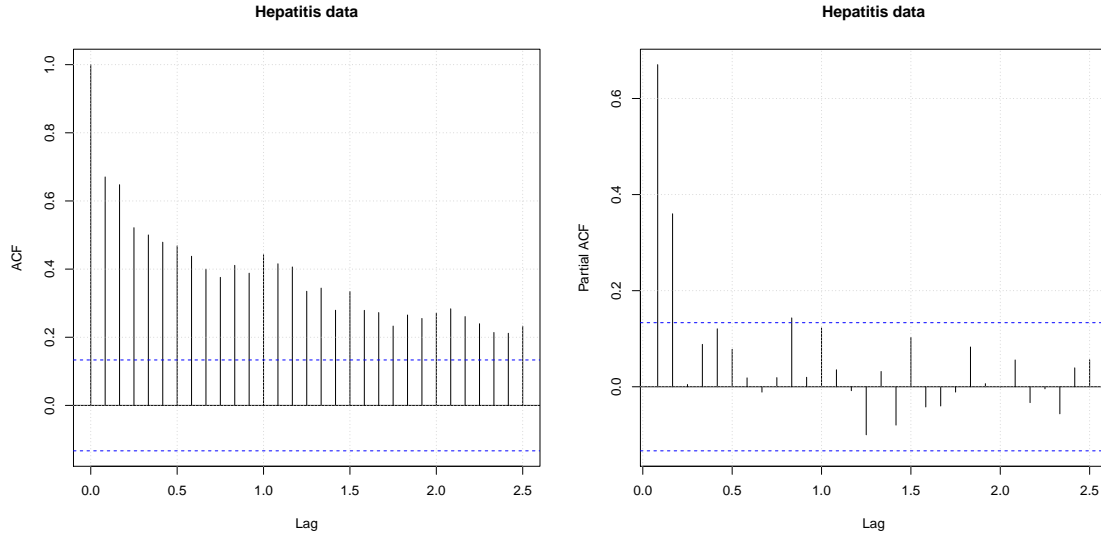


Figure 3: Plots of the estimated autocorrelation and partial autocorrelation functions for the hepatitis count time series data.

level at 5% based on the parametric bootstrap. The GLM/OLS estimates are also presented in this table. We observe a good agreement between the MCL and GLM/OLS methods for estimating β . On the other hand, this does not hold for the estimation of the parameters related to the latent process, which is in line with the simulated results presented in the paper. In this case, the composite likelihood estimation method yields better results for estimating the parameters of the latent process.

The analysis shows a significant decreasing linear trend ($\hat{\beta}_1 = -0.805$). Using a significance level at 10%, we find additionally that the semi-annual seasonality $\cos(2\pi t/6)$ and $\sin(2\pi t/6)$ are also significant with bootstrap intervals respectively equal to $(-0.109, -0.007)$ and $(0.002, 0.103)$ (note that these intervals have 90% confidence level while those in Table 5 have 95% confidence level). We check the performance of the MCL estimators based on the configuration estimated in this data analysis. From the parametric bootstrap with 1000 replicas used for getting the standard errors, we obtain the bootstrap-based estimates (averages of the estimates over all bootstrap samples) $\hat{\beta}_{boot} = (3.495, -0.692, -0.013, 0.015, -0.060, 0.053, -0.018, 0.012)^\top$, $\hat{\phi}_{boot} = 7.778$ and $\hat{\rho}_{boot} = 0.778$. So, we can observe a satisfactory performance of the MCL estimation method in this application since a good agreement between the MCL and bootstrap estimates.

We further investigate this data set by fitting the Poisson and negative binomial INGARCH(1,1) models with conditional mean $\lambda_t \equiv E(Y_t|Y_{t-1}, Y_{t-2}, \dots) = \beta_0 + \beta_1 t/216 + \beta_2 \cos(2\pi t/12) + \beta_3 \sin(2\pi t/12) +$

Table 5: Maximum composite likelihood estimates (MCL) of the parameters with their associated standard errors (S.E.) and parametric bootstrap-based confidence intervals (lower and upper bounds of the interval), with significance level at 5%, for the hepatitis data. The GLM/OLS estimates are also presented.

Parameters	MCL	GLM/OLS	S.E.	lower	upper
β_0 Intercept	3.575	3.625	0.146	3.233	3.779
β_1 Linear trend	-0.805	-0.912	0.259	-1.207	-0.215
β_2 $\cos(2\pi t/12)$	-0.011	-0.013	0.047	-0.114	0.078
β_3 $\sin(2\pi t/12)$	0.024	0.052	0.048	-0.081	0.114
β_4 $\cos(2\pi t/6)$	-0.063	-0.072	0.031	-0.119	0.000
β_5 $\sin(2\pi t/6)$	0.055	0.059	0.032	-0.007	0.117
β_6 $\cos(2\pi t/3)$	-0.021	-0.027	0.025	-0.072	0.027
β_7 $\sin(2\pi t/3)$	0.012	0.018	0.025	-0.037	0.062
ϕ dispersion	7.171	8.558	1.735	5.022	11.727
ρ correlation	0.786	0.618	0.069	0.629	0.894

$\beta_4 \cos(2\pi t/6) + \beta_5 \sin(2\pi t/6) + \beta_5 \cos(2\pi t/3) + \beta_6 \sin(2\pi t/3) + \gamma_1 Y_{t-1} + \kappa_1 \lambda_{t-1}$. We use the `tscount` package (Liboschik *et al.*, 2017) from the R program to fit the Poisson and negative binomial INGARCH models. The parameter estimates under the Poisson and NB INGARCH models were approximately the same given by $(\hat{\beta}, \hat{\gamma}_1, \hat{\kappa}_1) = (4.911, 7 \times 10^{-12}, 0.081, 2 \times 10^{-4}, 0.004, 0.176, 5 \times 10^{-10}, 0.025, 0.416, 0.391)$. The estimate of the overdispersion parameter (via the method of moments) of the negative binomial INGARCH model equals 0.088.

We now investigate the adequacy of the processes considered in this data analysis based on the PIT histograms, which are presented in Figure 4. We can observe from this figure that our NB model provides a suitable fit for the hepatitis data since all bars are close to 1, in contrast with the PIT histograms of the Poisson and negative binomial INGARCH(1,1) models.

We conclude this section by presenting the predictive values for the hepatitis data under the NBP. We consider the approach given in (4), and also those based on the surrogate predictive distribution (5) discussed in Section 3; we take the conditional expectation of the surrogate predictive distribution. In Figure 5, we present the hepatitis data with its associated predicted values based on the fitted NB process. From this figure, we see a good agreement between the empirical and predicted counts using the conditional expectation given in (4). The results obtained from the surrogate predictive distribution using all the past do not provide a satisfactory outcome. A possible explanation for this fact is that our inferential procedure considers only the pairwise likelihood functions and not the full likelihood function. Therefore, we recommend performing prediction based on the pairwise likelihood method such as (4) instead of the surrogate predictive approach

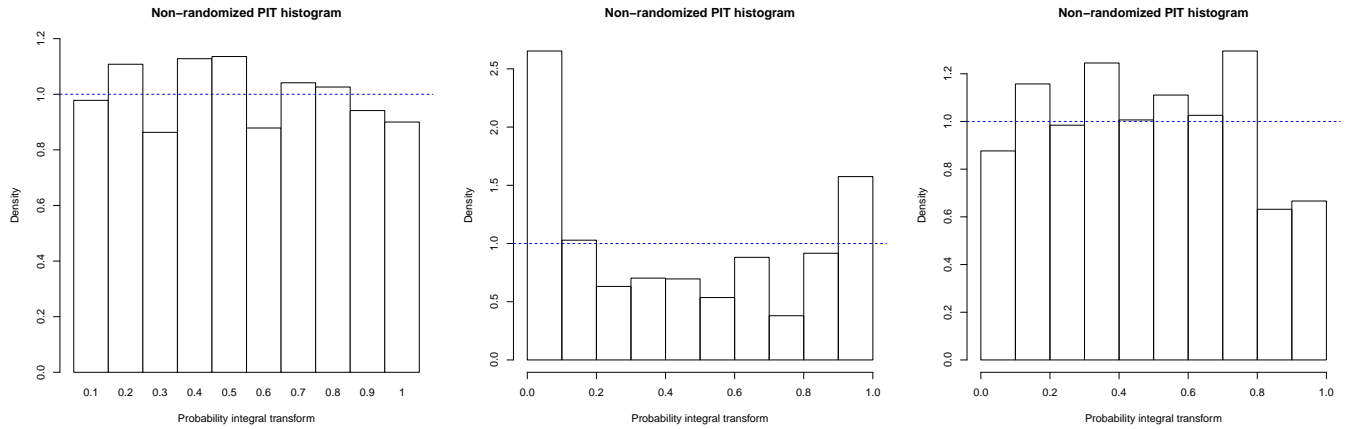


Figure 4: Probability integral transform plots under NBP (to the left), Poisson (in the middle) and negative binomial (to the right) INGARCH(1,1) models, for the hepatitis count time series.

using all the past (given in (5)) for this particular data set. On the other hand, we feel that more studies on this topic are necessary.

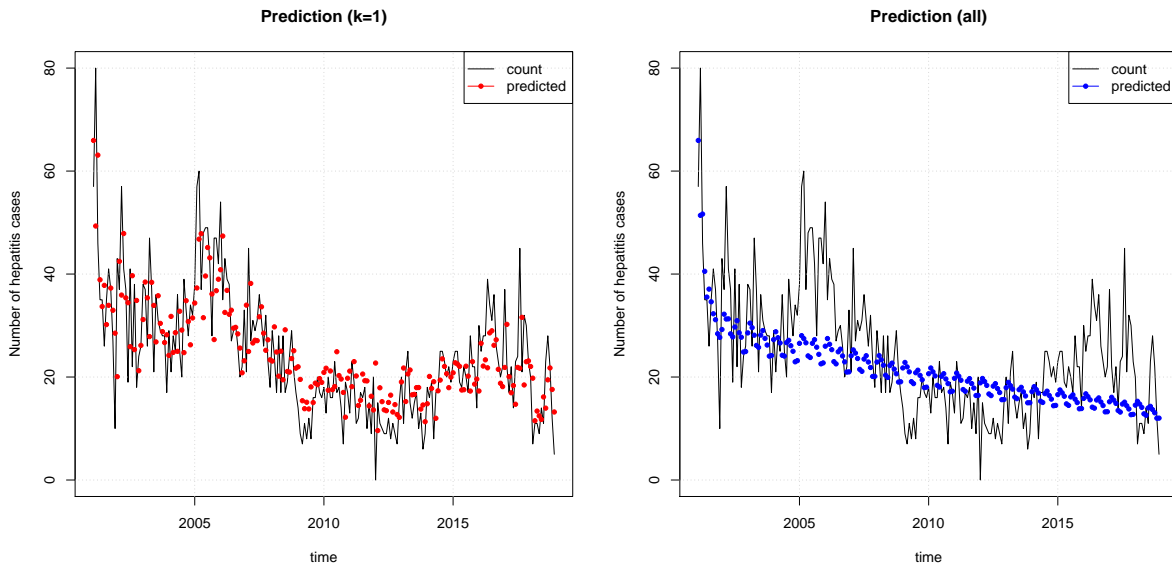


Figure 5: Hepatitis count data and its associated predicted values under the fitted NB process based on the conditional expectation (4) (to the left) and surrogate predictive distribution (5) (to the right).

6 Concluding remarks and future research

The main contribution of this paper is the proposal of a new count process with negative binomial marginals for overdispersed count time series data, which overcomes some drawbacks of existing models. We discussed inference based on a composite likelihood function and derived asymptotic properties such as strong consistency and asymptotic normality of the maximum composite likelihood estimators. The proposed methodology offers inference, prediction, and diagnostic to check model adequacy, for the modeling of count time series data. We applied the NBP to the monthly number of viral hepatitis cases in Goiânia (capital and largest city of the Brazilian state of Goiás) from January 2001 to December 2018. Our findings showed that the new approach is very promising for the modeling of seasonal disease count data.

A limitation of our model is that it can capture only positive autocorrelation. A possible way to extend it to include both positive and negative autocorrelation is to follow the approach recently proposed by [Jia *et al.* \(2018\)](#). More specifically, let $\{W_t\}_{t \in \mathbb{N}}$ be a stationary ARMA Gaussian model with zero mean and unit variance and define $Z_t = F^{-1}(\Phi(W_t))$, where $F(\cdot)$ and $\Phi(\cdot)$ are the cumulative distribution function of the gamma and standard normal distributions, respectively. With this, $\{Z_t\}_{t \in \mathbb{N}}$ is a stationary gamma process with a flexible autocorrelation structure. This latent process can be used for defining a negative binomial model $\{Y_t\}_{t \in \mathbb{N}}$ able for both positive and negative autocorrelations. On the other hand, this model is not tractable in terms of composite likelihood as the negative binomial process proposed in this paper.

We now discuss possible high-order extensions of our NB process. A first attempt to do this is to generalize the GAR(1) autoregression allowing higher-order $p \geq 2$ in the negative binomial model. Another way is to consider the latent process following a Generalized Autoregressive Moving Average Models (GARMA) model based on the gamma distribution; see [Benjamin *et al.* \(2003\)](#). Under this last assumption, the marginals of the resulting process are no longer negative binomial distributed. We call the attention that AR(1) models are commonly used as latent processes rather than AR(p) models with $p \geq 2$, as discussed in the introduction of this paper. Furthermore, such extensions can suffer from non-identifiability problems, as discussed by [Alzahrani *et al.* \(2018\)](#). The authors showed that the [Zeger \(1988\)](#) model with latent AR(p) Gaussian process can suffer from lack of identifiability in the AR parameters so giving support for the AR(1) model. This can be the case of a possible high-order extension of our NB process, but it is beyond the scope of the current paper.

The prediction methods discussed in the paper are inspired by the composite likelihood approach, which is our main focus. As pointed out by the Associate Editor, there is a range of other approaches for state-space models like the sequential Monte Carlo which could be addressed.

A last point of interest is to propose a multivariate negative binomial process for dealing with correlated time series count data. This is an ongoing work and we hope to report our findings soon. As suggested by one of the referees, the particle MCMC algorithm by [Alzahrani *et al.* \(2018\)](#) could be an alternative to the composite likelihood approach. We feel that this strategy can be valuable for parameter estimation of the multivariate extension that we are working in since the composite likelihood approach, in that case, is cumbersome.

Items in the Supplementary Material

Additional results, R code, and dataset: This online Supplementary Material file contains two sections. The first section gives additional simulated results from Section 4 and the second one provides an R code for fitting the NB process under the estimation methods discussed in Section 3. The count time series data analyzed in the paper is available in the Supplementary Material.

Acknowledgements

We would like to thank the Associate Editor and three anonymous Referees for their careful reading and suggestions which leads to a substantial improvement of this paper. W. Barreto-Souza and H. Ombao would like to acknowledge support by KAUST Research Fund and NIH 1R01EB028753-01. W. Barreto-Souza also thanks the financial support from *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq-Brazil, grant number 305543/2018-0).

References

- AL-OSH, M.A. & ALY, E.-E.A.A. (1992). First order autoregressive time series with negative binomial and geometric marginals. *Comm. Statist. Theory Methods*. **21**, 2483–2492.
- AL-OSH, M.A. & ALZAID, A.A. (1987). First-order integer valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **8**, 261–275.

- ALY, E.E.A.A. & BOUZAR, N. (1994). Explicit stationary distributions for some Galton-Watson processes with immigration. *Stoch. Models.* **10**, 499–517.
- ALZAID, A.A. & AL-OSH, M.A. (1988). First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties. *Stat. Neerl.* **42**, 53–61.
- ALZHRANI, N., NEAL, P., SPENCER, S., MCKINLEY, T. & TOULOUPOU, P. (2018). Model selection for time series of count data. *Comput. Stat. Data Anal.* **122**, 33–44.
- BARRETO-SOUZA, W. (2015). Zero-modified geometric INAR(1) process for modelling count time series with deflation or inflation of zeros. *J. Time Ser. Anal.* **36**, 839–852.
- BENJAMIN, M.A., RIGBY, R.A. & STASINOPOULOS, D.M. (2003). Generalized autoregressive moving average models. *J. Am. Stat. Assoc.* **98**, 214–223.
- BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307–327.
- BRÄNNÄS, K. & JOHANSSON, P. (1994). Time series count data regression. *Commun. Stat. Theory Methods.* **23**, 2907–2925.
- BROTO, C. & RUIZ, E. (2004). Estimation methods for stochastic volatility models: A survey. *J. Econ. Surv.* **18**, 613–650.
- CHAN, K.S. & LEDOLTER, J. (1995). Monte Carlo EM estimation for time series models involving counts. *J. Am. Stat. Assoc.* **90**, 242–252.
- CHERNICK, M.R. (1981). On strong mixing and Leadbetter’s D condition. *J. Appl. Prob.* **18**, 764–769.
- CHRISTOU, V. & FOKIANOS, K. (2014). Quasi-likelihood inference for negative binomial time series models. *J. Time Ser. Anal.* **35**, 55–78.
- CHRISTOU, V. & FOKIANOS, K. (2015). Estimation and testing linearity for non-linear mixed Poisson autoregressions. *Eletron. J. Stat.* **9**, 1357–1377.
- COX, D.R. (1981). Statistical analysis of time series. *Scand. J. Stat.* **8**, 93–115.
- CUI, Y. & LUND, R. (2009). A new look at time series of counts. *Biometrika.* **96**, 781–792.
- CZADO, C., GNEITING, T. & HELD, L. (2009). Predictive model assessment for count data. *Biometrics.* **65**, 1254–1261.
- DAVIS, R.A., DUNSMUIR, W.T.M. & WANG, Y. (1999). Modeling time series of count data. In *Asymptotics, Nonparametrics and Time Series*, Ed. S. Gosh, pp. 63–114. New York: Marcel Dekker.
- DAVIS, R.A., DUNSMUIR, W.T.M. & STRETT, S. (2003). Observation-driven models for Poisson counts. *Biometrika.* **90**, 777–790.

- DAVIS, R.A., DUNSMUIR, W.T.M. & WANG, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*. **87**, 491–505.
- DAVIS, R.A. & WU, R. (2009). A negative binomial model for time series of counts. *Biometrika*. **96**, 735–749.
- DAVIS, R.A. & LIU, H. (2016). Theory and inference for a class of nonlinear models with applications to time series of counts. *Stat. Sin.* **26**, 1673–1707.
- EFRON, B. & TIBSHARANI, R. (1993). *An Introduction to Bootstrap*. New York: Ed. Chapman & Hall.
- ENCISO-MORA, V., NEAL, P.J. & SUBBA RAO, T. (2009a). Efficient order selection algorithms for integer valued ARMA processes. *J. Time Series. Anal.* **30**, 1–18.
- ENCISO-MORA, V., NEAL, P.J. & SUBBA RAO, T. (2009b). Integer valued AR processes with explanatory variables. *Sankhyā, Ser. B.* **71**, 248–263.
- FERLAND, R., LATOUR, A. & ORAICHI, D. (2006). Integer-valued GARCH process. *J. Time Ser. Anal.* **27**, 923–942.
- FREELAND, R.K. & MCCABE, B.P.M. (2004). Analysis of low count time series data by Poisson autoregression. *J. Time Ser. Anal.* **25**, 701–722.
- FOKIANOS, K. & FRIED, R. (2010). Interventions in INGARCH processes. *J. Time Ser. Anal.* **31**, 210–225.
- FOKIANOS, K., RAHBEK, A. & TJØSTHEIM, D. (2009). Poisson autoregression. *J. Am. Stat. Assoc.* **104**, 1430–1439.
- FOKIANOS, K. (2011). Some recent progress in count time series. *Statistics* **45**, 49–58.
- FOKIANOS, K. & FRIED, R. (2012). Interventions in log-linear Poisson autoregression. *Stat. Model.* **12**, 299–322.
- GODAMBE, V.P. (1960). An optimum property of regular maximum likelihood equation. *Ann. Math. Stat.* **31**, 1208–1211.
- HYNDMAN, R.J. & KOEHLER, A.B. (2006). Another look at measures of forecast accuracy. *Int. J. Forecast.* **22**, 679–688.
- JACOBS, P. & LEWIS, P. (1983). Stationary discrete autoregressive-moving average time series generated by mixtures. *J. Time Ser. Anal.* **4**, 19–36.
- JIA, Y., KECHAGIAS, S., LIVSEY, J., LUND, R. & PIPIRAS, V. (2018). Latent Gaussian count time series modeling. arXiv: 1811.00203. 1–38.
- KEDEM, B. & FOKIANOS, K. (2002). *Regression Models for Time Series Analysis*. Wiley: New Jersey.
- KIM, S., SHEPHARD, N. & CHIB, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econ. Stud.* **65**, 361–393.

- LIBOSCHIK, T., FOKIANOS, K. & FRIED, R. (2017). `tscount`: An R package for analysis of count time series following generalized linear models. *J. Stat. Softw.* **82**, 1–51.
- MCKENZIE, E. (1985). Some simple models for discrete variate time series. *Water Resour. Bull.* **21**, 645–650.
- MCKENZIE, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Adv. Appl. Probab.* **18**, 679–705.
- NELDER, J.A. & MEAD, R. (1965). A simplex method for function minimization. *Comput. J.* **7**, 308–313.
- NG, C.T., JOE, H., KARLIS, D. & LIU, J. (2011). Composite likelihood for time series models with a latent autoregressive process. *Stat. Sinica* **21**, 279–305.
- PEDELI, X. & VARIN, C. (2020). Pairwise likelihood estimation of latent autoregressive count models. *Stat. Methods Med. Res.* **29**, 3278–3293.
- QUDDUS, M.A. (2008). Time series count data models: an empirical application to traffic accidents. *Accid. Anal. Prev.* **40**, 1732–1741.
- R DEVELOPMENT CORE TEAM (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, <http://www.R-project.org>.
- RISTIĆ, M., BAKOUCH, H.S. & NASTIĆ, A.S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *J. Stat. Plan. Inference.* **139**, 2218–2226.
- SCOTTO, M.G., WEISS, C.H. & GOUVEIA, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Stat. Model.* **15**, 590–618.
- SILVA, R.B. & BARRETO-SOUZA, W. (2019). Flexible and robust mixed Poisson INGARCH models. *J. Time Ser. Anal.* **40**, 788–814.
- SIM, C.H. (1990). First-order autoregressive models for gamma and exponential processes. *J. Appl. Probab.* **27**, 325–332.
- STEUTEL, F.W. & VAN HARN, K. (1990). Discrete analogues of self-decomposability and stability. *Ann. Probab.* **7**, 893–899.
- VARIN, C. (2008). On composite marginal likelihoods. *AStA Adv. Stat. Anal.* **92**, 1–28.
- VARIN, C., REID, N. & FIRTH, D. (2011). An overview of composite likelihood methods. *Stat. Sin.* **21**, 5–42.
- VIDONI, P. (2018). A note on predictive densities based on composite likelihood methods. *Metron.* **76**, 31–48.
- ZEGER, S.L. (1988). A regression model for time series of counts. *Biometrika.* **75**, 621–629.
- ZHU, F. (2011). A negative binomial integer-valued GARCH model. *J. Time Ser. Anal.* **32**, 34–67.

ZHU, R. & JOE, H. (2010). Negative binomial time series models based on expectation thinning operators. *J. Statist. Plann. Inference* **140**, 1874–1888.

WEISS, C. (2018). *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons Ltd.

WEISS, C. (2020). Stationary count time series models. *WIREs Comput Stat.* doi.org/10.1002/wics.1502

Appendix

Appendix A1: Proofs

Proof of Proposition 2.6. We have that the conditional distribution of Y_{t+j} and Y_t given $(Z_{t+j}, Z_t) = (z_{t+j}, z_t)$ are independent Poisson distributions with mean $\mu_{t+j}z_{t+j}$ and $\mu_t z_t$, respectively, where (Z_{t+j}, Z_t) is a bivariate gamma random vector with joint probability function given in (1). With this fact, we have that

$$\begin{aligned} \Pr(Y_{t+j} = y, Y_t = x) &= \int_0^\infty \int_0^\infty \Pr(Y_{t+j} = y, Y_t = x | Z_{t+j} = z_{t+j}, Z_t = z_t) f(z_{t+j}, z_t) dz_{t+j} dz_t \\ &= \frac{\mu_{t+j}^y \mu_t^x}{x! y!} \int_0^\infty \int_0^\infty \exp(-\mu_{t+j} z_{t+j}) \exp(-\mu_t z_t) z_{t+j}^y z_t^x f(z_{t+j}, z_t) dz_{t+j} dz_t, \end{aligned}$$

where $f(z_{t+j}, z_t)$ is given in (1). After some algebra, we find

$$\begin{aligned} \Pr(Y_{t+j} = y, Y_t = x) &= \frac{\phi^{\phi+1} \mu_t^x \mu_{t+j}^y}{(1 - \rho^j) \Gamma(\phi) \rho^{j(\phi-1)/2} y! x!} \int_0^\infty z_t^{\frac{\phi-1}{2} + x} e^{-z_t(\mu_t + \frac{\phi}{1-\rho^j})} \times \\ &\quad \left\{ \int_0^\infty z_{t+j}^{\frac{\phi-1}{2} + y} e^{-z_{t+j}(\mu_{t+j} + \frac{\phi}{1-\rho^j})} I_{\phi-1} \left(2\phi \frac{\sqrt{\rho^j z_{t+j} z_t}}{1 - \rho^j} \right) dz_{t+j} \right\} dz_t. \end{aligned} \tag{8}$$

By using the series representation of the Bessel function and the Fubini's Theorem (or the Monotone Convergence Theorem), we can interchange the signs of integral and sum (related to the series

representation of the Bessel function) in the above integral inside the brackets which is given by

$$\begin{aligned} & \int_0^\infty z_{t+j}^{\frac{\phi-1}{2}+y} e^{-z_{t+j}(\mu_{t+j}+\frac{\phi}{1-\rho^j})} I_{\phi-1} \left(2\phi \frac{\sqrt{\rho^j z_{t+j} z_t}}{1-\rho^j} \right) dz_{t+j} = \\ & \sum_{k=0}^\infty \frac{\left(\phi \frac{\sqrt{\rho^j z_t}}{1-\rho^j} \right)^{2k+\phi-1}}{\Gamma(k+\phi)k!} \int_0^\infty z_{t+j}^{\phi+y+k-1} e^{-z_{t+j}(\mu_{t+j}+\frac{\phi}{1-\rho^j})} dz_{t+j} = \\ & \sum_{k=0}^\infty \frac{\left(\phi \frac{\sqrt{\rho^j}}{1-\rho^j} \right)^{2k+\phi-1}}{\Gamma(k+\phi)k!} \frac{z_t^{k+\frac{\phi-1}{2}} \Gamma(\phi+y+k)}{\left(\mu_{t+j} + \frac{\phi}{1-\rho^j} \right)^{\phi+y+k}}. \end{aligned}$$

Using the above result in (8), we applied again the Fubini's Theorem and obtain similarly as above the following expression for the joint probability function

$$\Pr(Y_{t+j} = y, Y_t = x) = \frac{\phi^{2\phi}(1-\rho^j)^{\phi+x+y} \mu_t^x \mu_{t+j}^y \sum_{k=0}^\infty \lambda_{tj}^k \frac{\Gamma(\phi+x+k)\Gamma(\phi+y+k)}{\Gamma(\phi+k)k!}}{\Gamma(\phi)x!y! [(1-\rho^j)\mu_t + \phi]^{x+\phi} [(1-\rho^j)\mu_{t+j} + \phi]^{y+\phi}},$$

where $\lambda_{tj} = \frac{\phi^2 \rho^j}{[(1-\rho^j)\mu_t + \phi][(1-\rho^j)\mu_{t+j} + \phi]}$. Since $|\lambda_{tj}| < 1$, the above infinite sum can be expressed by $\frac{\Gamma(x+\phi)\Gamma(y+\phi)}{\Gamma(\phi)} {}_2F_1(x+\phi, y+\phi; \phi, \lambda_{tj})$, which completes the proof. \square

Proof of Proposition 2.7. We have that $E(Y_{t+j}|Y_t) = E[E(Y_{t+j}|Y_t, Z_t)|Y_t] = E[E(Y_{t+j}|Z_t)|Y_t]$, where the last equality follows from the fact that Y_{t+j} and Y_t are conditionally independent given Z_t . Using properties of the GAR process $\{Z_t\}_{t \in \mathbb{N}}$ and the conditional expectation (2), we obtain that the conditional expectation of Y_{t+j} given Z_t is $E(Y_{t+j}|Z_t) = \mu_{t+j}E(Z_{t+j}|Z_t) = \mu_{t+j}[1 + \rho^j(Z_t - 1)]$. Hence, it follows that $E(Y_{t+j}|Y_t) = \mu_{t+j}[1 + \rho^j(E(Z_t|Y_t) - 1)] = \mu_{t+j} \left(1 + \rho^j \frac{Y_t - \mu_t}{\mu_t + \phi} \right)$, where the last equality follows from the fact that $Z_t|Y_t = y \sim G(y + \phi, \mu + \phi)$. In a similar fashion, we obtain the conditional expectation $E(Y_{t+j}^2|Y_t)$ and hence compute the conditional variance $\text{Var}(Y_{t+j}|Y_t)$. \square

Appendix A2: Functions for generating and fitting the NBP

```
# Example: Generating and fitting a NBP trajectory

n<-200 # sample size
t<-seq(1:n)
x<-rbind(array(1,c(1,n)),cos(2*pi*t/6)) # covariates
# Parameter values
beta<-c(1,0.5)
phi<-0.5
rho<-0.7
eta<-t(x)%*%beta # linear predictor
mu<-exp(eta) # mean function of the NB process

# Fixing the seed for the reproducibility of the results
set.seed(2020)

# Generation of a NB process trajectory
y<-r.nbp(n,rho,phi,mu)

# Fitting the simulated NBP trajectory
z<-rbind(y,x)
fit.model<-fit.nbp(z)

# CL estimates of beta, phi and rho (in this order)
fit.model$CL_est
0.807 0.466 0.432 0.671

# Standard errors of the CL estimates
fit.model$CL_se
0.245 0.137 0.082 0.094

# OLS estimates of beta, phi and rho (in this order)
fit.model$OLS_est
0.805 0.457 0.346 0.638

# CLS estimates of beta, phi and rho (in this order)
fit.model$CLS_est
0.750 0.585 0.049 0.563
```