Conway-Maxwell-Poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data

Abstract. In this work, we propose a dynamic regression model based on the Conway-Maxwell-Poisson (CMP) distribution with time-varying conditional mean depending on covariates and lagged observations. This new class of Conway-Maxwell-Poisson autoregressive moving average (CMP-ARMA) models is suitable for the analysis of time series of counts. The CMP distribution is a two-parameter generalization of the Poisson distribution that allows the modeling of underdispersed, equidispersed, and overdispersed data. Our main contribution is to combine this dispersion flexibility with the inclusion of lagged terms to model the conditional mean response, inducing an autocorrelation structure, usually relevant in time series. We present the conditional maximum likelihood estimation, hypothesis testing inference, diagnostic analysis, and forecasting along with their asymptotic properties. In particular, we provide closed-form expressions for the conditional score vector and conditional Fisher information matrix. We conduct a Monte Carlo experiment to evaluate the performance of the estimators in finite sample sizes. Finally, we illustrate the usefulness of the proposed model by exploring two empirical applications.

Keywords: CMP-ARMA; Conway-Maxwell-Poisson distribution; Time series of counts; Forecasts; Overdispersion; Underdispersion

1. Introduction

Models for time series of counts have received considerable and growing attention in recent decades. These series are commonly observed in real-world applications such as economics (Freeland and McCabe, 2004), medicine (Franke and Seligmann, 1993), and epidemiology (Zeger and Qaqish, 1988).

An appropriate and flexible approach for count time series is to apply the generalized autoregressive moving average model (GARMA) proposed by Benjamin et al. (2003). The GARMA model combines the autoregressive moving average model (ARMA) (Box et al., 2015) with the generalized linear model (GLM) methodology (McCullagh and Nelder, 1989), enabling the inclusion of autoregressive and moving average components. This model can be applied in the analysis of count data observed over time using the conditional Poisson, Negative Binomial, or binomial distributions.

The most popular distribution for modeling count data is the Poisson distribution (Shmueli et al., 2005). However, in practice, this distribution is not always suitable since many real data do not adhere to the assumption of equidispersion

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jtsa.12550

(where the mean and variance are equal). Often the data are overdispersed (variance is greater than the mean), this phenomenon has received considerable attention in the literature (MacDonald and Bhamani, 2018). The phenomenon of underdispersion (variance is lower than the mean) occurs less frequently, and the choice of distributions is restricted (Zhu, 2012).

In recent years, the modeling of count time series with overdispersion and underdispersion has received great attention, and one of the distributions that accommodates these disperson cases is the Conway-Maxwell-Poisson (CMP) distribution. This distribution has been applied in several fields, including marketing, transportation, and epidemiology (Sellers et al., 2012). In time series settings, Zhu (2012) proposed an integer-valued generalized autoregressive conditional heteroscedastic model (INGARCH) with CMP distribution. Mamode Khan et al. (2018) introduced an observation-driven integer-valued moving average model of order 1 (INMA(1)) with CMP innovations under non-stationary moment conditions. Despite this last model includes the thinning operator considering the serial correlation, it is more appropriate for low counts. Moreover, MacDonald and Bhamani (2018) introduced the class of stationary hidden Markov models with CMP distribution as state-dependent distribution. Although the models proposed by Zhu (2012) and MacDonald and Bhamani (2018) are able to model underdispersion and overdispersion, they do not include covariates. The model proposed by Mamode Khan et al. (2018) includes regressors, but the mean is not directly modeled, leading to a complicate interpretation of parameters. For the proposed model in this paper, the mean of the conditional distribution is directly modeled, making the model parameters easily interpretable.

The present paper introduces a dynamic regression model for time series following a CMP distribution. In order to define the proposed model, we shall follow similar construction as the GARMA model (Benjamin et al., 2003). The proposed CMP-ARMA model can be used for modeling time series of counts with equidispersion, underdispersion, and overdispersion. One of the advantages of time series models based on GLM is that they straightforwardly describe covariate effects and negative autocorrelations (Liboschik et al., 2017). In addition to the GARMA model, several time series models based on GLM with different distributions have been considered in the literature (Li, 1991, 1994; Fokianos and Kedem, 2004; Rocha and Cribari-Neto, 2009; Bayer et al., 2017).

Our chief goal is to propose a new class of dynamic regression time series models for non-negative discrete data, with equidispersion, overdispersion, and/or underdispersion, based on the structure developed in Benjamin et al. (2003). We use the parameterization of the CMP distribution in terms of its mean as proposed by Huang (2017). For this purpose, we present the main properties of the model, the conditional maximum likelihood estimators (CMLE), and some residual and diagnostic tools. In addition, we provide a Monte Carlo study to evaluate the CMLE performance and stationarity conditions.

This paper is organized as follows. Section 2 reviews the CMP distribution and its reparametrization, where the mean of the distribution is rewritten as a function of the original parameters. Section 3 introduces the proposed model. In Section 4, we develop the estimation and inference for the new CMP-ARMA model based on the conditional maximum likelihood theory, including closed-form expressions for the conditional score vector and conditional Fisher information matrix. We also discuss the construction of confidence intervals and hypothesis testing. Section 5 discusses some diagnostic measures and forecasting. In Section 6, we present the results of the Monte Carlo simulation study. Section 7 illustrates the flexibility of the proposed model through two empirical applications. Finally, some conclusions are given in Section 8. Details on the derivation of the conditional score vector and conditional Fisher information matrix are presented in the Appendices.

2. The CMP distribution

Although Poisson models are popularly known for modeling count data, many real data sets usually do not present equidispersion as in the Poisson distribution. Another widely used distribution is the Negative Binomial, which can capture overdispersion. However, analyzing underdispersed counts is a big challenge. Recently, Shmueli et al. (2005) suggested the use of the CMP distribution to model equidispersed, overdispersed, and underdispersed counts, which was originally developed by Conway and Maxwell (1962) as a model for queuing systems with state-dependent service times.

Let Y be a random variable with CMP(λ, ν) distribution, then its probability mass function is given by

$$Pr(Y = y \mid \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)}, \quad y = 0, 1, 2, \ldots, \lambda > 0, \nu \geq 0,$$

where $Z(\lambda, \nu) = \sum_{s=0}^{\infty} \lambda^s/(s!)^\nu$ is a normalization constant, and ν is the dispersion parameter such that $\nu > 1$ represents underdispersion and $0 \leq \nu < 1$ overdispersion.

The CMP distribution generalizes the Poisson distribution by relaxing the assumption of linearity of the ratio of consecutive probabilities, such that

$$\frac{Pr(Y = y - 1)}{Pr(Y = y)} = \frac{y^\nu}{\lambda}.$$

This generalization allows heavier or lighter tails compared to the Poisson distribution (Sellers and Shmueli, 2010). One of the advantages of the CMP distribution is that, in addition to the Poisson distribution ($\nu = 1$), we have the Geometric ($\nu = 0; \lambda < 1$) and the Bernoulli ($\nu \to \infty$ with probability $\lambda/(1 + \lambda)$) distributions as particular cases.

The moments of the CMP distribution are expressed using the following recursive method

$$E(Y^{r+1}) = \begin{cases} \lambda E(Y + 1)^{1-\nu}, & r = 0 \\ \lambda \frac{d}{d\lambda} E(Y^r) + E(Y)E(Y^r), & r > 0. \end{cases}$$

Since the above method does not have a closed-form solution, an asymptotic approximation for $Z(\lambda, \nu)$ can be used. Shmueli et al. (2005) presented an approximate form for the moments of the distribution given by

$$E(Y) \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu} \text{ and } V(Y) \approx \frac{\lambda^{1/\nu}}{\nu} \qquad (1)$$

which is particularly accurate for $\nu \leq 1$ or $\lambda > 10^{\nu}$.

Sellers and Shmueli (2010) proposed the CMP regression model using the original parameterization $\log(\lambda_i) = x_i^T \beta$, where the mean is modeled through the approximation in Equation (1). Note that this approximation is accurate for $\nu \leq 1$ or $\lambda > 10^{\nu}$, and the mean is indirectly modeled. In order to avoid such issues, Huang (2017) proposed a reparametrization of the CMP model, where instead of using an approximation for the mean value, the mean of the counts $\mu = E(Y)$ is modeled directly assuming that $\log(\mu_i) = x_i^T \beta$, as defined in generalized linear models. This reparametrization makes the model simpler and easily interpretable.

The reparameterized CMP distribution (Huang, 2017), denoted by CMP_{μ}, is given by the probability mass function

$$Pr(Y = y \mid \mu, \nu) = \frac{\lambda(\mu, \nu)^y}{(y!)^{\nu} Z(\lambda(\mu, \nu), \nu)}, \qquad \mu \geq 0;\ \nu \geq 0;\ y = 0, 1, 2, \ldots, \qquad (2)$$

where $\lambda(\mu, \nu)$ is a function of μ and ν, given by the solution for

$$0 = \sum_{s=0}^{\infty} (s - \mu) \frac{\lambda^s}{(s!)^{\nu}}.$$

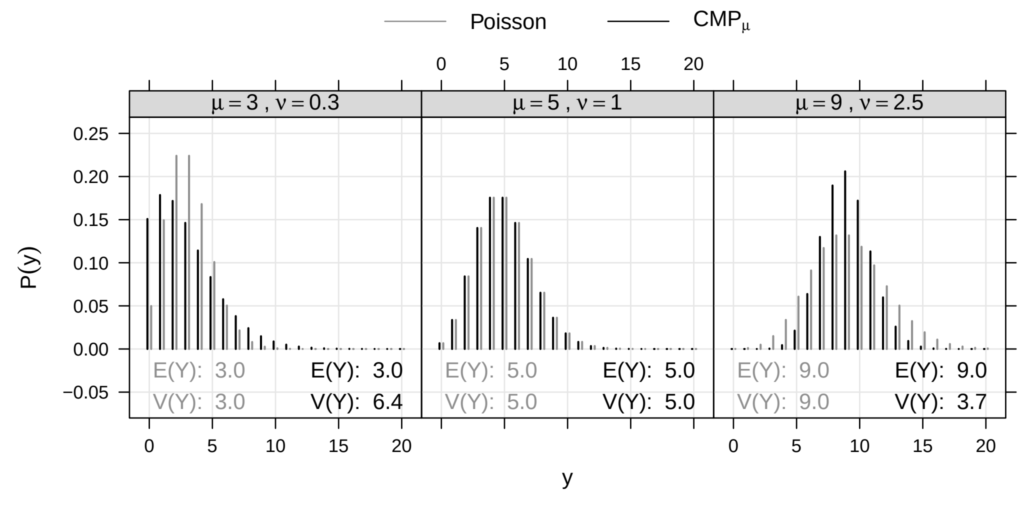


Fig. 1. Shapes of the Poisson and CMP_{μ} probability mass functions for different values of μ and ν

.

In Figure 1, some CMP_μ distributions are depicted for different mean values μ and dispersion parameter ν and compared to Poisson distributions of same mean μ. The left panel shows the distribution of overdispersion counts ($\nu < 1$). In the central panel, the CMP_μ with $\nu = 1$ corresponds to the Poisson distribution. Finally, the right panel exposes a case in which the distribution presents underdispersion ($\nu > 1$). Note that a large dispersion parameter value ν condenses the distribution around the mean, and a small dispersion parameter value extends the distribution away from the mean.

The CMP_μ distribution can be expressed as the exponential family of two parameters (Huang, 2017). This is analogous to the result obtained by Shmueli et al. (2005) for the standard CMP distribution. For fixed ν, the CMP_μ distribution belongs to the one parameter exponential family.

3. The proposed model

In this work, we propose a new dynamic regression model for random variables with a CMP distribution observed over time. To define the model, we include an ARMA time series structure in the conditional mean to accommodate the presence of serial correlation. Since for time series analysis it is convenient to work with the mean response, we use the CMP reparametrization given in Equation (2).

Let $Y = (Y_1, \ldots, Y_n)^\top$ be a vector of n random variables, where the conditional distribution of Y_t , $t = 1, \ldots, n$, given the previous information set $\mathcal{F}_{t-1} = \{Y_{t-1}, \ldots, Y_1; \mu_{t-1}, \ldots, \mu_1\}$, follows a CMP distribution with mean parameter μ_t and dispersion parameter ν. The conditional probability function of Y_t, given $\mathcal{F}_{t-1}$, is defined as

$$Pr(Y_t = y_t \mid \mathcal{F}_{t-1}, \mu_t, \nu) = \frac{\lambda(\mu_t, \nu)^{y_t}}{(y_t!)^\nu Z(\lambda(\mu_t, \nu), \nu)}, \quad y = 0, 1, 2, \ldots, \tag{3}$$

where $\lambda(\mu_t, \nu)$ is a function of μ_t and ν, given by the solution for

$$0 = \sum_{s=0}^{\infty} (s - \mu_t) \frac{\lambda^s}{(s!)^\nu}, \tag{4}$$

and $Z(\lambda(\mu_t, \nu), \nu) = \sum_{s=0}^{\infty} \lambda(\mu_t, \nu)^s/(s!)^\nu$ is a normalization funtion, and the conditional mean of Y_t is given by $E(Y_t \mid \mathcal{F}_{t-1}) = \mu_t$. As in Huang (2017), the variance does not have a closed-form expression. However, the main advantage of the reparametrization proposed by Huang (2017) is its ability to model the mean directly, making it possible to compare the CMP-ARMA model with the Poisson and Negative Binomial GARMA models.

As in the CMP regression model (Huang, 2017), the conditional mean μ_t is related to the linear predictor η_t by a twice-differentiable one-to-one monotonic function $g(\cdot)$, called the link function. However, unlike the CMP regression model, the linear predictor of the proposed model has an additional component, allowing for autoregressive and moving average terms to be included as

$$\eta_t = g(\mu_t) = \alpha + \boldsymbol{x}_t^\top \boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j [g(y_{t-j}) - \boldsymbol{x}_{t-j}^\top \boldsymbol{\beta}] + \sum_{j=1}^{q} \theta_j r_{t-j}, \tag{5}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)^\top$ is the r-dimensional unknown parameter vector, $\boldsymbol{x}_t = (x_1, \ldots, x_r)^\top$ is the r-dimensional explanatory variables vector, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$ are the autoregressive and moving average coefficients, respectively, r_t is a random error, and α is an intercept. In this paper, we consider the errors measured on the predictor scale $r_t = g(y_t) - g(\mu_t)$ as in Rocha and Cribari-Neto (2009), Bayer et al. (2017), and Benjamin et al. (2003).

The proposed CMP-ARMA(p, q) model is defined by (3) and (5). Due to the restriction $\mu_t \geq 0$, we choose the logarithm as link function because it provides non-negative values for $\mu_t = g^{-1}(\eta_t)$ regardless the values assigned to η_t. In order to allow the use of the logarithm link function with modeling count series containing observations equal to zero, we replace y_{t-j} in Equation (5) for $y_{t-j}^* = max(y_{t-j}, c)$, with threshold c such that $0 < c < 1$. This procedure allows replacing $y_{t-j} = 0$ by an arbitrary small value c. Note that the dynamic part of Equation (5) is the same as in Benjamin et al. (2003).

The CMP-ARMA model contains the GARMA models with the Poisson, Geometric, and Bernoulli distributions as special cases. (see Appendix A for details).

4. Parameter estimation

Let $y_1, \ldots, y_t$, $t = 1, \ldots, n$, be a sample from a CMP-ARMA(p, q) model. Let $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top, \nu)^\top$ be the regression parameter vector. Based, conditionally, on the m first observations, where $m = \max(p,q)$, the conditional log-likelihood function is given by

$$\ell(\boldsymbol{\gamma}) = \sum_{t=m+1}^{n} \log f(y_t | \mathcal{F}_{t-1}) = \sum_{t=m}^{n} \ell_t(\mu_t, \nu), \tag{6}$$

where

$$\ell_t(\mu_t, \nu) = y_t \log(\lambda(\mu_t, \nu)) - \nu \log(y_t!) - \log Z(\lambda(\mu_t, \nu), \nu).$$

4.1. Conditional score vector

The conditional score vector $U(\boldsymbol{\gamma}) = (U_\alpha(\boldsymbol{\gamma}), \boldsymbol{U}_\beta(\boldsymbol{\gamma})^\top, \boldsymbol{U}_\phi(\boldsymbol{\gamma})^\top, \boldsymbol{U}_\theta(\boldsymbol{\gamma})^\top, U_\nu(\boldsymbol{\gamma}))^\top$ is obtained by taking the first derivative of the conditional log-likelihood function with respect to each element of $\boldsymbol{\gamma}$ and is expressed in matrix form as

$$
\begin{aligned}
U_\alpha(\boldsymbol{\gamma}) &= \boldsymbol{\delta}^\top \boldsymbol{T}\boldsymbol{V}(\boldsymbol{y}-\boldsymbol{\mu}),\\
\boldsymbol{U}_\beta(\boldsymbol{\gamma}) &= \boldsymbol{M}^\top \boldsymbol{T}\boldsymbol{V}(\boldsymbol{y}-\boldsymbol{\mu}),\\
\boldsymbol{U}_\phi(\boldsymbol{\gamma}) &= \boldsymbol{P}^\top \boldsymbol{T}\boldsymbol{V}(\boldsymbol{y}-\boldsymbol{\mu}),\\
\boldsymbol{U}_\theta(\boldsymbol{\gamma}) &= \boldsymbol{Q}^\top \boldsymbol{T}\boldsymbol{V}(\boldsymbol{y}-\boldsymbol{\mu}),\\
U_\nu(\boldsymbol{\gamma}) &= \sum_{t=m+1}^{n} E_{\mu_t,\nu}\left[\log(y_t!)(\mu_t-y_t)\right]\frac{(y_t-\mu_t)}{V(\mu_t,\nu)} - \left[\log(y_t!) - E_{\mu,\nu}\log(y_t!)\right],
\end{aligned} \tag{7}
$$

where $\boldsymbol{y}=(y_{m+1},\ldots,y_n)^\top$, $\boldsymbol{\mu}=(\mu_{m+1},\ldots,\mu_n)^\top$, $\boldsymbol{T}=\operatorname{diag}\left\{1/g'(\mu_{m+1}),\ldots,1/g'(\mu_n)\right\}$, $\boldsymbol{\delta}=\left(\frac{\partial\eta_{m+1}}{\partial\alpha},\ldots,\frac{\partial\eta_n}{\partial\alpha}\right)^\top$, $\boldsymbol{V}=\operatorname{diag}\left\{1/V(\mu_{m+1},\nu),\ldots,1/V(\mu_n,\nu)\right\}$, $\boldsymbol{M}$ is an $(n-m)\times r$ matrix whose (i,j)-th element is given by $\dfrac{\partial\eta_i}{\partial\beta_j}$, $\boldsymbol{P}$ is an $(n-m)\times p$ matrix whose (i,j)-th element is equal to $\dfrac{\partial\eta_i}{\partial\phi_j}$, $\boldsymbol{Q}$ is an $(n-m)\times q$ matrix where the (i,j)-th element is $\dfrac{\partial\eta_i}{\partial\theta_j}$, and $V(\mu_t,\nu)=\sum\limits_{y=0}^{\infty}\dfrac{(y-\mu_t)^2\lambda(\mu_t,\nu)^y}{(y!)^\nu Z(\lambda(\mu_t,\nu),\nu)}$. To compute the derivative of η_t with respect to the unknown parameters, the error in (5) is defined by $r_t=g(y_t)-g(\mu_t)$, then

$$
\begin{aligned}
\frac{\partial\eta_t}{\partial\alpha} &= 1+\sum_{j=1}^{q}\theta_j\frac{\partial r_{t-j}}{\partial\alpha} = 1-\sum_{j=1}^{q}\theta_j\frac{\partial\eta_{t-j}}{\partial\alpha},\\
\frac{\partial\eta_t}{\partial\beta_i} &= x_{ti}-\sum_{j=1}^{p}\phi_j x_{(t-j)i}-\sum_{j=1}^{q}\theta_j\frac{\partial\eta_{t-j}}{\partial\beta_i},\\
\frac{\partial\eta_t}{\partial\phi_i} &= g(y_{t-i})-\boldsymbol{x}_{\boldsymbol{t-i}}^\top\boldsymbol{\beta}-\sum_{j=1}^{q}\theta_j\frac{\partial\eta_{t-j}}{\partial\phi_i},\\
\frac{\partial\eta_t}{\partial\theta_i} &= (g(y_{t-i})-\eta_{t-i})-\sum_{j=1}^{q}\theta_j\frac{\partial\eta_{t-j}}{\partial\theta_i},
\end{aligned}
$$

(see Appendix B for details).

Notice that recursions are required only when the model includes moving average components. In this case, it is necessary to choose initial values for η_t and its derivatives. Here, we assume $\eta_t=g(y_t)$, and the initial values for the derivatives equal zero, both for $t=1,2,\ldots,m$. See Benjamin et al. (1998) for details.

The solution of the estimation equation $\boldsymbol{U}(\boldsymbol{\gamma})=\boldsymbol{0}$, where $\boldsymbol{0}$ is the null vector in $\mathbb{R}^{r+p+q+2}$, provides the CMLE of $\boldsymbol{\gamma}$, denoted by $\widehat{\boldsymbol{\gamma}}$. This system does not have an analytical solution, being necessary the use of iterative numerical methods to obtain an approximate solution. Here, we use the `nlminb` optimization function available in `R`, which is a reverse-communication trust-region quasi-Newton method from the **Port** library (Gay, 1990).

4.2. Conditional information matrix

In this section, we derive the conditional Fisher information matrix, obtained by taking partial derivatives of second order of the conditional log-likelihood function given in (6). Let $W = \mathrm{diag}\{w_1, \ldots, w_n\}$, where

$$w_t = -\frac{[E_{\mu_t,\nu}\left(\log(y_t!)(y_t - \mu_t)\right)]^2}{V(\mu_t,\nu)} + Var_{\mu_t,\nu}(\log(y_t)).$$

The conditional Fisher information matrix is given by

$$\boldsymbol{K}(\gamma) = \begin{pmatrix} \boldsymbol{\delta}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{\delta} & \boldsymbol{\delta}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{M} & \boldsymbol{\delta}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{P} & \boldsymbol{\delta}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{Q} & \boldsymbol{0} \\ (\boldsymbol{\delta}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{M})^\top & \boldsymbol{M}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{M} & \boldsymbol{M}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{P} & \boldsymbol{M}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{Q} & \boldsymbol{0} \\ (\boldsymbol{\delta}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{P})^\top & (\boldsymbol{M}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{P})^\top & \boldsymbol{P}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{P} & \boldsymbol{P}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{Q} & \boldsymbol{0} \\ (\boldsymbol{\delta}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{Q})^\top & (\boldsymbol{M}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{Q})^\top & (\boldsymbol{P}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{Q})^\top & \boldsymbol{Q}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \mathrm{tr}(\boldsymbol{W}) \end{pmatrix}$$

(see Appendix C for details).

Under regularity conditions, the maximum likelihood estimators are consistent and asymptotically normally distributed (Andersen, 1970). Thus, when the sample size is sufficiently large, the CMLE $\widehat{\boldsymbol{\gamma}}$ of the parameter vector $\boldsymbol{\gamma}$ has an approximately normal distribution with mean $\boldsymbol{\gamma}$ and variance-covariance matrix $\boldsymbol{K}^{-1}$, that is,

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\phi}} \\ \widehat{\boldsymbol{\theta}} \\ \widehat{\nu} \end{pmatrix} \sim N_{r+p+q+2}\left(\begin{pmatrix} \alpha \\ \boldsymbol{\beta} \\ \boldsymbol{\phi} \\ \boldsymbol{\theta} \\ \nu \end{pmatrix}, \boldsymbol{K}^{-1} \right), \tag{8}$$

where $N_{r+p+q+2}$ denotes the $(r+p+q+2)$-dimensional normal distribution and $\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}$, and $\widehat{\nu}$ the CMLE of $\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}$, and ν, respectively.

4.3. Confidence intervals and hypothesis testing

Let γ_i be the i-th element of the parameter vector $\boldsymbol{\gamma}$ and k^{ii} the i-th diagonal element of $(\boldsymbol{K}(\boldsymbol{\gamma}))^{-1}$. From (8), the asymptotic distribution is

$$\frac{\widehat{\gamma}_i - \gamma_i}{\sqrt{k^{ii}}} \sim N(0{,}1).$$

Therefore, an $100(1-\alpha)\%$ asymptotic confidence interval for each parameter γ_i, $i = 1, \ldots, (r+p+q+2)$, is given by

$$\left[\widehat{\gamma}_i - z_{1-\alpha/2}\sqrt{k^{ii}}; \widehat{\gamma}_i + z_{1-\alpha/2}\sqrt{k^{ii}}\right],$$

where $\Phi(z_{1-\alpha/2}) = 1 - \alpha/2$, with $\Phi(\cdot)$ being the cumulative density function of the standardized normal distribution $N(0,1)$.

Consider the following null hypothesis $\mathcal{H}_0 : \gamma_i = \gamma_i^0$, where γ_i^0 is a specified value for the unknown parameter γ_i, versus the alternative hypothesis $\mathcal{H}_1 : \gamma_i \neq$

γ_i^0. A convenient statistic to test individual parameters is the so-called z statistic (Pawitan, 2001) given by

$$Z = \frac{\widehat{\gamma}_i - \gamma_i}{\sqrt{k^{ii}}}. \tag{9}$$

This statistic is based on the signed square root of Wald's statistic. Under $\mathcal{H}_0$ and large sample sizes, the limiting null distribution of z is standard normal.

It is also possible to perform general hypothesis testing inference using the log-partial likelihood ratio, Wald, and score statistics. Under $\mathcal{H}_0$, all of these test statistics converge to a χ^2 distribution. See Kedem and Fokianos (2005) for further details.

The z statistic in Equation (9) can be used to test equidispersion in the data. We considered the following hypotheses

$$\mathcal{H}_0 : \nu = 1 \text{ versus } \mathcal{H}_1 : \nu \neq 1.$$

Here, the null hypothesis is that the data present equidispersion and the alternative hypothesis is that the data are over/underdispersed. The non-rejection of the null hypothesis indicates that the use of the Poisson model is appropriate.

5. Model diagnosis and forecasting

In this section, we present some model selection criteria as well as some procedures to test the adequacy and goodness-of-fit of the proposed model. We also provide a method for out-of-sample forecasting.

5.1. Deviance

One way to measure goodness-of-fit is by means of scaled deviance, which is defined as twice the difference between the conditional log-likelihood function of the saturated (where each μ_t is estimated directly from y_t) and fitted models, that is

$$D = 2(\tilde{\ell} - \widehat{\ell}) = 2\left(\sum_{t=m+1}^{n} \ell_t(y_t, \widehat{\nu}) - \sum_{t=m+1}^{n} \ell_t(\widehat{\mu}_t, \widehat{\nu})\right).$$

When the fitted model is correct, the test statistic D is approximately χ^2 distributed with $n - (r + p + q + m + 2)$ degrees of freedom (Benjamin et al., 2003; Fokianos and Kedem, 2004).

5.2. Model Selection Criteria

For comparison and selection among several competing models, we can use model selection criteria available in the literature. The basic idea is to select a parsimonious model, in other words, a model that is well-fitted and has a small/sufficient number of parameters. Two widely used model selection criteria are the Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion

(BIC) (Schwarz et al., 1978), given, respectively, by

$$AIC = -2\widehat{\ell} + 2(p+q+r+2),$$
$$BIC = -2\widehat{\ell} + \log(n)(p+q+r+2).$$

Among the considered models, the one with smaller value of AIC and/or BIC is selected.

5.3. Residual Analysis

Residual analysis is important to check if all model assumptions are valid (Kedem and Fokianos, 2005), and also used to identify poorly fitted observations, i.e., observations not well explained by the model (Feng et al., 2017). For the proposed CMP-ARMA model, we consider the quantile residual (Dunn and Smyth, 1996) as Pearson and deviance residuals are non-normally distributed in count data (Benjamin et al., 2003). Let $a_t = F(y_t - 1 \mid \mathcal{F}_{t-1})$ and $b_t = F(y_t \mid \mathcal{F}_{t-1})$, where F is the fitted conditional cumulative distribution function. For the discrete distribution function, the randomized quantile residual for y_t is defined by

$$r_t^{(q)} = \Phi^{-1}(u_t),$$

where Φ^{-1} is the quantile function of the standard normal distribution, and u_i is a random variable that is uniformly distributed on $(a_t, b_t]$. If the model fitted to the data is correctly specified, these residuals should be independent and normally distributed, with zero mean and unit variance.

5.4. PIT Histograms

Another diagnostic tool for model assessment is the Probability Integral Transform (PIT), which follows a uniform distribution if the fitted model is correctly specified (Jung and Tremayne, 2011; Jung et al., 2015). Although the PIT applies to continuous distributions, Czado et al. (2009) proposed a non-randomized yet uniform version of the PIT as an alternative to time series models for count data. The conditional cumulative distribution function of observed counts y_t is

$$F^{(t)}(u \mid \mathcal{F}_{t-1}) = \begin{cases} 0, & u \leq F(y_t - 1 \mid \mathcal{F}_{t-1}), \\ \dfrac{u - F(y_t - 1 \mid \mathcal{F}_{t-1})}{F(y_t \mid \mathcal{F}_{t-1}) - F(y_t - 1 \mid \mathcal{F}_{t-1})}, & F(y_t - 1 \mid \mathcal{F}_{t-1}) \leq u \leq F(y_t \mid \mathcal{F}_{t-1}), \\ 1, & u \geq F(y_t \mid \mathcal{F}_{t-1}), \end{cases}$$

where $F(y_t \mid \mathcal{F}_{t-1}) = \sum_{y=0}^{y_t} P(y \mid \mathcal{F}_{t-1}, \widehat{\mu}_t, \widehat{\nu})$ with $P(y \mid \mathcal{F}_{t-1}, \mu, \nu)$ defined in (3).

The assessment of the fitted model can be carried out by comparing the mean PIT, defined as

$$\overline{F}(u) = (n-m)^{-1} \sum_{t=m+1}^{n} F^{(t)}(u \mid \mathcal{F}_{t-1}), \qquad 0 \leq u \leq 1,$$

with the cumulative distribution function of a standard uniform random variable.

To perform this comparison, Czado et al. (2009) proposed plotting a non-randomized PIT histogram with J equally spaced bins, where the height f_j for bin $j = 1, \ldots, J$ is computed by

$$f_j = \overline{F}(j/J) - \overline{F}((j-1)/J).$$

Czado et al. (2009) suggest $J = 10$ or $J = 20$ as good choices for the number of bins in the PIT histogram.

5.5. Forecasting

Consider the problem of forecasting a value for the observed response h steps ahead, $h \in \mathbb{N}$, denoted by $\widehat{y}_{n+h} = \widehat{y}_n(h)$. Forecasts of future observations for the CMP-ARMA model can be obtained by applying the CMLE in (5),

$$\widehat{y}_n(h) = \exp\left(\widehat{\alpha} + \boldsymbol{x}_{n+h}^\top \widehat{\boldsymbol{\beta}} + \sum_{j=1}^{p} \widehat{\phi}_j [g(y_{n+h-j}) - \boldsymbol{x}_{n+h-i}^\top \widehat{\boldsymbol{\beta}}] + \sum_{j=1}^{q} \widehat{\theta}_j \widehat{r}_{n+h-j}\right),$$

where

$$[g(y_t)] = \begin{cases} g(\widehat{\mu}_t), & t > n, \\ g(y_t), & t \leq n. \end{cases}$$

It is also possible to generate confidence intervals for the forecasts using quantiles.

6. Monte Carlo simulation study

In this section, we present a Monte Carlo simulation study to evaluate the asymptotic properties of the CMLE for the CMP-ARMA model and investigate the stationarity conditions.

6.1. Asymptotic properties

We consider samples from a CMP-ARMA(1,1) model and three different values for the dispersion parameter: $\nu \in \{0.5, 1.0, 2.0\}$. Thus, we have overdispersion, equidispersion, and underdispersion, respectively. The CMP-ARMA(1,1) model with systematic component is given by

$$\log(\mu_t) = \alpha + \beta_1 x_t + \phi_1[\log(y_{t-1}) - \beta_1 x_{t-1}] + \theta_1 r_{t-1}, \qquad t = 2, \ldots, n,$$

where $x_t = sin(2\pi t/12)$, $\alpha = 1.5$, $\beta_1 = 0.5$, $\phi = 0.5$ and $\theta = 0.3$. All routines were implemented in the R statistical computing environment (R Core Team, 2019) and are available upon request. All results are based on 5,000 replications of each combination for the sample sizes $n = 50, 100, 200, 400$. We evaluate mean, percentage relative bias (RB %), defined as $\{E(\widehat{\theta}) - \theta\}/\theta$, and mean squared error (MSE).

Table 1 shows that the overall performance of the CMLE improves as the sample size increases. We observe that the estimator of β_1 is nearly unbiased for small sample sizes, for example $n = 50$. We also note that the MSE decreases as the sample size increases, indicating consistency of the CMLE. Moreover, the moving average parameter θ is overestimated while the autoregressive parameter ϕ is underestimated in all scenarios.

6.2. Stationarity conditions

Benjamin et al. (2003) derived stationarity conditions with marginal mean and variance of the dependent variable Y_t in the GARMA model with identity link function for some exponential family distributions. As for the GARMA model, the stationarity for the CMP-ARMA model requires the invertibility of the polynomial $\Phi(B) = 1 - \phi_1 B - \ldots, \phi_p B^p$. When the link function is different from identity, the parameter restrictions to ensure stationarity appear to be intractable.

In order to investigate the region of the parameter space for which a CMP-ARMA process is stationary, we carried out a Monte Carlo simulation study similar to that presented in Benjamin et al. (2003). We simulated 1,000 realizations of length 200 of a CMP-ARMA(1,1) model with logarithmic link function, threshold $c = 0.1$, and intercept $\alpha = \ln(2)$ in combination with the following parameter values $\phi, \theta \in \{-0.4, 0, 0.4, 0.8\}$ and $\nu \in \{0.5, 1.0, 2.0\}$. As in Benjamin et al. (2003), we compared the empirical distributions at times 150, 175, and 200 using a chi-square goodness-of-fit test to check for nonstationarity in each parameter combination. We found no evidence of nonstationarity for any of the parameter combinations considered.

7. Empirical applications

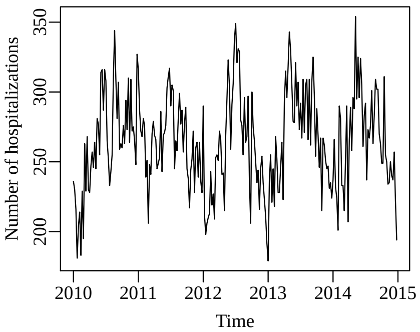
In this section, we present and discuss two empirical applications to show the applicability of the proposed model. We also compare the CMP-ARMA (proposed) model with the Poisson and Negative Binomial GARMA models, considering the parameterization of the Negative Binomial distribution used in Evans (1953). Using this parameterization, the conditional mean and the conditional variance of Y_t given $\mathcal{F}_{t-1}$ are $E(Y_t \mid \mathcal{F}_{t-1}) = \mu_t$ and $Var(Y_t \mid \mathcal{F}_{t-1}) = (\sigma + 1)\mu_t$, respectively, σ being the dispersion parameter. It is noteworthy that for a Poisson distribution we have $E(Y_t \mid \mathcal{F}_{t-1}) = Var(Y_t \mid \mathcal{F}_{t-1}) = \mu_t$. The Poisson and Negative Binomial GARMA models were fitted using the *garmaFit* function from *gamlss.util* (Stasinopoulos and Rigby, 2016) library in the `R` software.

7.1. Overdispersion data: weekly number of hospitalizations

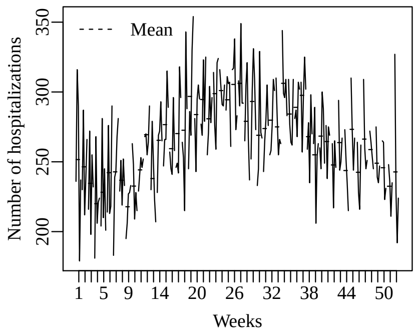
According to the United Nations Population Division (UN, 2015), the number of people with age over 60 is expected to grow 56% worldwide between 2015 and 2030. In São Paulo, the largest city in Brazil, with 11 million inhabitants in 2010 (IBGE, 2011), 11% of its inhabitants belong to this age range. The number of admissions for respiratory problems is supposed to increase overtime for elderly people (Alencar, 2018). Given its relevance, understanding and modeling the behavior of the number of hospitalizations due to respiratory diseases is necessary, as well as evaluating the out-of-sample forecasts. This helps the State to take preventive actions regarding public health, for example, to plan the vaccination calendar.

The data consist of the weekly number of hospitalizations due to respiratory diseases for people aged over 60 years in the city of São Paulo-Brazil from January 2010 to December 2014, yielding a sample size of $n = 260$. The first 250 observations were used to model the time series, and the remaining 10 observations were used to evaluate the out-of-sample forecasts. These data were obtained from the Hospitalization Information System of the Ministry of Health (available at Datasus website `http://datasus.saude.gov.br/`).

The empirical mean and variance of the data are 264.61 and 1201.60, respectively, indicating that the data are overdispersed. The original series and its seasonal component are plotted in Figures 2(a) and 2(b), respectively. The data present seasonal behavior since the mean number of hospitalizations increases in the winter (June to September) and decreases during the summer. We modeled the seasonal effect through sine and cosine functions with annual cycle. However, the estimated coefficient of the sine function was close to zero so that we considered only the cosine function as covariate in the model: $x_t = \cos(2\pi t/52))$, $t = 1, \ldots, n$. In addition, we chose the logarithm as link function. First, we fitted the CMP-GLM model to the data, but their residual autocorrelation (ACF) and partial autocorrelation function (PACF), as shown in Figures 3(a) and 3(b), respectively, indicate a second-order autoregressive autocorrelation structure. Hence, we fitted the proposed model to the data.



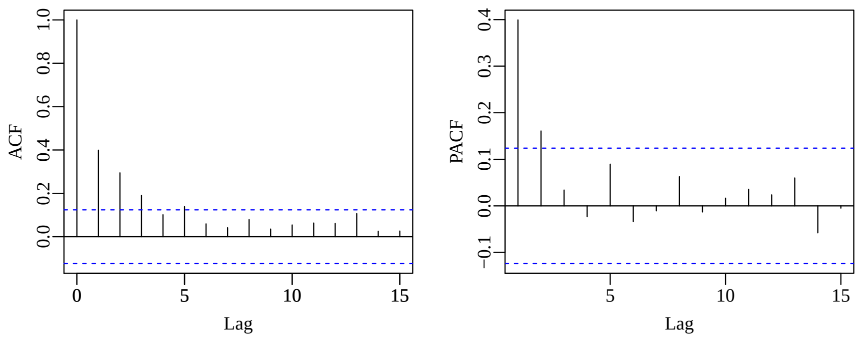
(a) Observed values



(b) Seasonality

Fig. 2. (a) Observed number of hospitalizations and (b) seasonal component present in the data.

We considered different orders (p,q) to fit the proposed model, and we selected the CMP-ARMA(2,0) model as it presented the lowest AIC and/or BIC. The chosen model corroborates with the structure indicated by the residuals of the GLM model. Table 2 shows the parameter estimates and corresponding standard errors (SE), z statistics, p-values and information criteria. For comparison purposes, Table 2 also shows the fitted Poisson and Negative Binomial GARMA$(2, 0)$ models, which were



(a) Residual ACF (b) Residual PACF

Fig. 3. (a) ACF and (b) PACF of randomized quantile residuals for the CMP-GLM model.

the best Poisson and Negative Binomial GARMA models for this data set. Note that the CMP-ARMA and Negative Binomial GARMA models present similar parameter estimates and information criteria. It is noteworthy that these results are expected since the estimated dispersion parameters indicate that the data are overdispersed ($\nu = 0.4035$ and $\sigma = 1.471$). We test $\mathcal{H}_0 : \nu = 1$ versus $\mathcal{H}_1 : \nu \neq 1$ and rejected the null hypothesis with p-value < 0.0001. We note that the AIC and BIC values of the Poisson GARMA model are higher than those of the other two models.

Diagnostic plots for the CMP-ARMA model are presented in Figure 4. Figures 4(a) and 4(b) display the ACF and PACF of randomized quantile residuals, respectively. These plots along with the Box-Ljung statistic (Ljung and Box, 1978) (using 15 lags) indicate that the residuals are not autocorrelated (p-value $= 0.935$). The sequence of residuals in Figure 4(c) seems to be oscillating around zero with constant variance. Figure 4(d) presents the empirical and normal quantiles. The analysis of these two plots indicates that the residuals are approximately normally distributed. Figure 4(e) shows the non-randomized PIT histogram of the fitted model with $J = 10$. The uniformity of the PIT (Figure 4(e)) suggests that the CMP-ARMA(2,0) is a suitable model to the data, and Figure 6(a) shows that the model provides a good fit. Diagnostic plots for the fitted Negative Binomial GARMA(2,0) model exhibit similar results and are omitted for brevity. Figure 5 presents diagnostic plots for the Poisson GARMA model. Although the ACF (Figure 5(a)) and PACF (Figure 5(b)) of randomized quantile residuals indicate that residual autocorrelations are not significant, the sequence of residuals (Figure 5(c)) and normal probability (Figure 5(d)) plots indicate that the residuals are non-normally distributed. Furthermore, the PIT histogram (Figures 5(e)) and the normal probability plots suggest that the Poisson GARMA model is unable to capture overdispersion in the data.

Finally, the out-of-sample forecasts using the fitted CMP-ARMA model are presented in Figure 6(b) along with the observed values for comparison reason. The mean absolute percentage error (MAPE) between the observed data (y_{n+h}) and out-of-sample forecasts ($\widehat{y}_{n+h}$), for $h = 1, \ldots, 10$, is 5.08%.

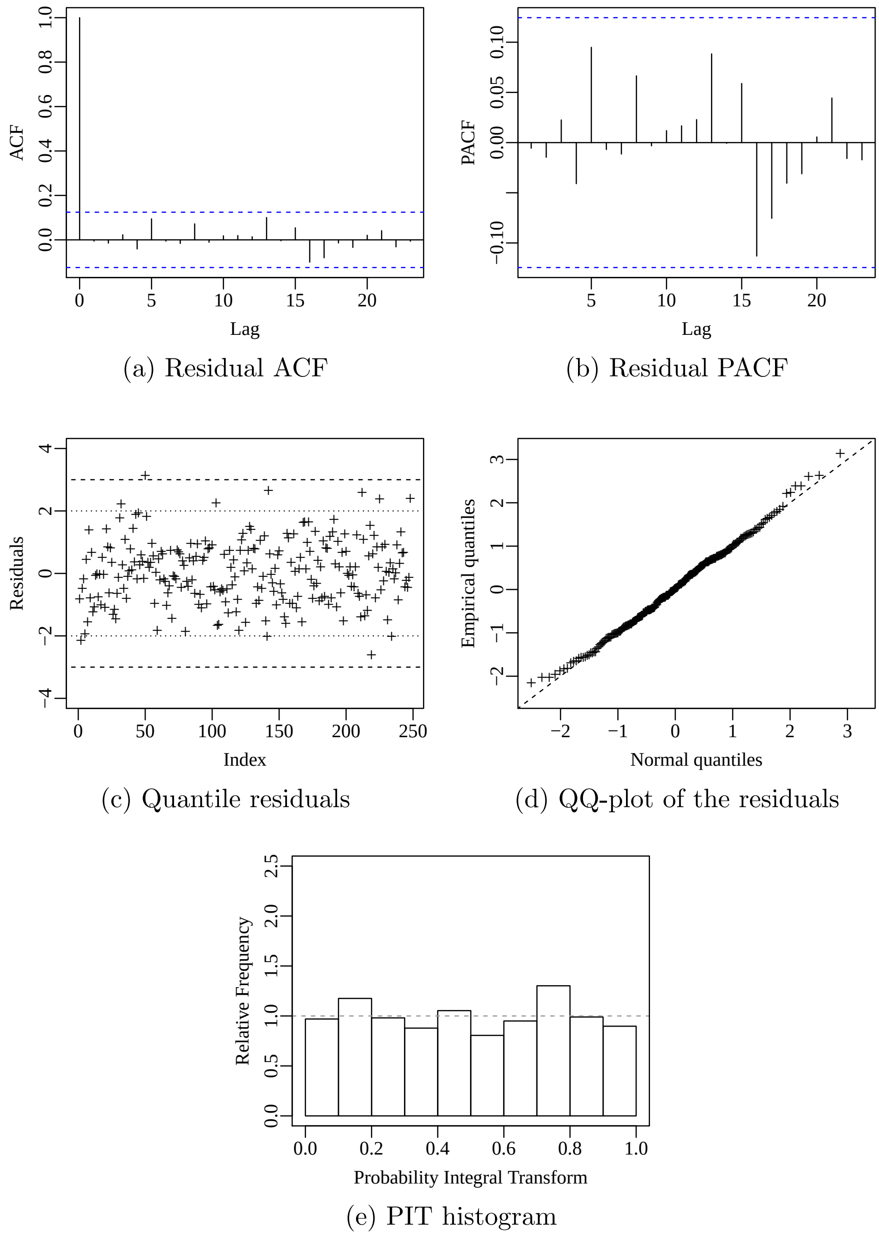
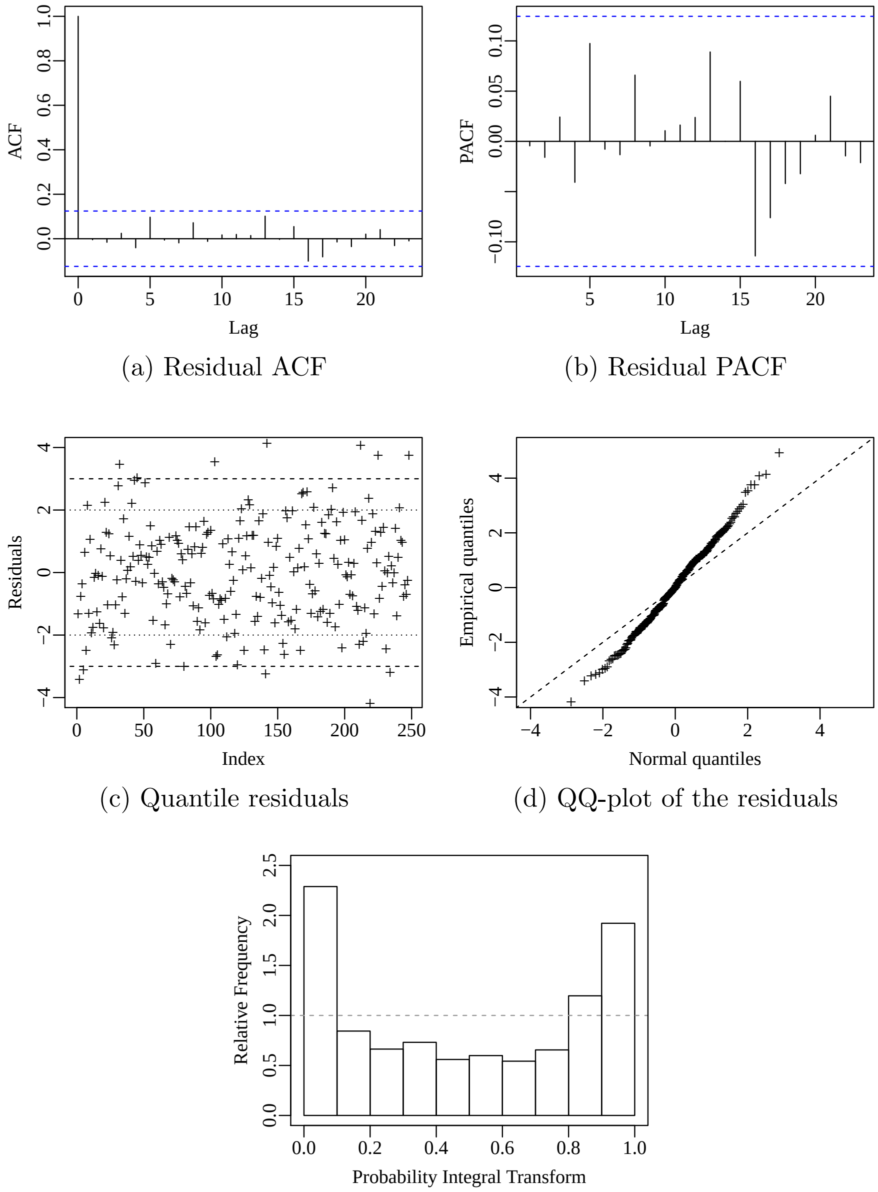


Fig. 4. Diagnostic plots for the fitted CMP-ARMA model; weekly number of hospitalizations data.

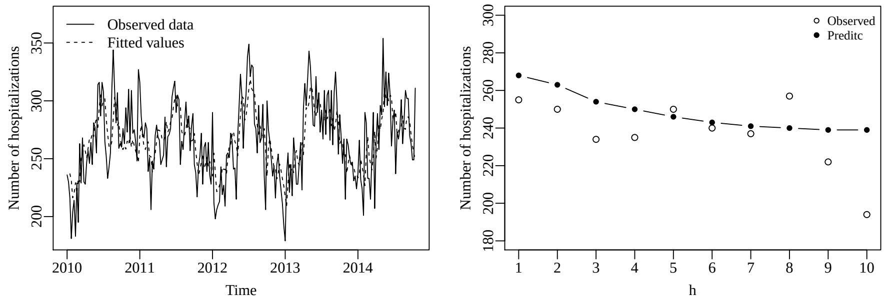


(e) PIT histogram

Fig. 5. Diagnostic plots for the fitted Poisson-GARMA model; weekly number of hospitalizations data.

7.2. Underdispersed data: pedestrian counts

We analyzed the second data set to illustrate the flexibility of the proposed model for underdispersed data. The data set consists of 505 counts of the number of



(a) Fitted values (b) Forecasts

Fig. 6. (a) Fitted values and (b) forecasts; weekly number of hospitalizations data.

pedestrians traversing a city block observed at 5-second intervals. These data were originally presented by Fürth (1918) and later analyzed by, among others, Jung and Tremayne (2006) and MacDonald and Bhamani (2018). The sample mean and variance are 1.592 and 1.508, respectively, indicating underdispersion in the data. Jung and Tremayne (2006) analyzed the present data by fitting some first and second order integer-valued autoregressive (INAR) and integer-valued moving average (INMA) models. However, they concluded that none of these models fit the data satisfactorily. Recently, MacDonald and Bhamani (2018) modeled these data using a class of stationary hidden Markov models with CMP distributions as state-dependent distributions.

Figure 7 presents the count time series. Initially, we fitted the CMP-GLM model with logarithm link function to the data. The ACF and PACF of randomized quantile residuals are displayed in Figures 8(a) and 8(b), respectively, exhibiting a significant serial dependence structure. Thus, CMP-ARMA, Poisson GARMA, and Negative Binomial GARMA models were fitted to the data. Based on the information criteria, the CMP-ARMA(1,1), Negative Binomial GARMA(1,1), and the Poisson GARMA(1,1) models were selected.

Table 3 presents the parameter estimates and corresponding SE, p-values, and information criteria for the three models. Note that the estimated dispersion parameter for the CMP-ARMA model is $\widehat{\nu} = 2.4428$, indicating existence of underdispersion in the data and inappropriateness of the GARMA models with Poisson and Negative Binomial conditional distributions. Note that the test $\mathcal{H}_0 : \nu = 1$ versus $\mathcal{H}_1 : \nu \neq 1$ rejected the null hypothesis (p-value < 0.0001). Also, the AIC and BIC values favor the CMP-ARMA model as they are smaller than those of the Poisson and Negative Binomial GARMA models.

Figure 9 shows diagnostic plots for the CMP-ARMA model. The ACF and PACF of randomized quantile residuals shown in Figures 9(a) and 9(b), respectively, along with Box-Ljung statistic based on 15 lags, confirm the assumption that there is no significant residual autocorrelation (p-value $= 0.146$). Figures 9(c) and 9(d) show the residual and normal probability plots, respectively. Both plots indicate that the

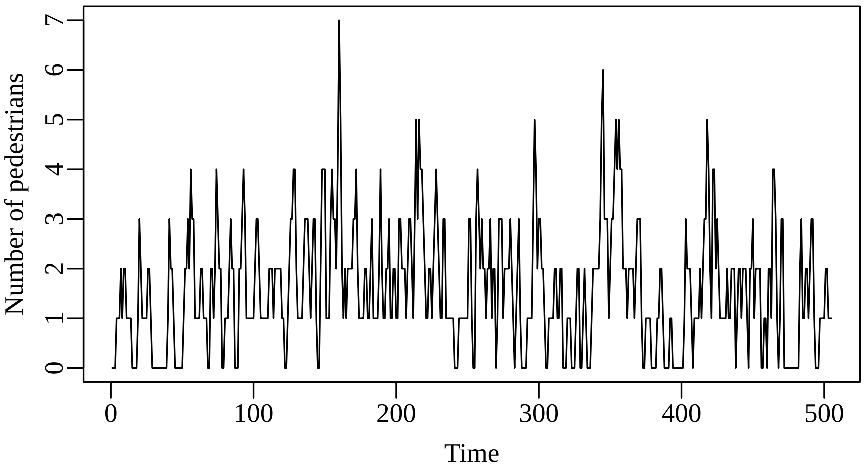
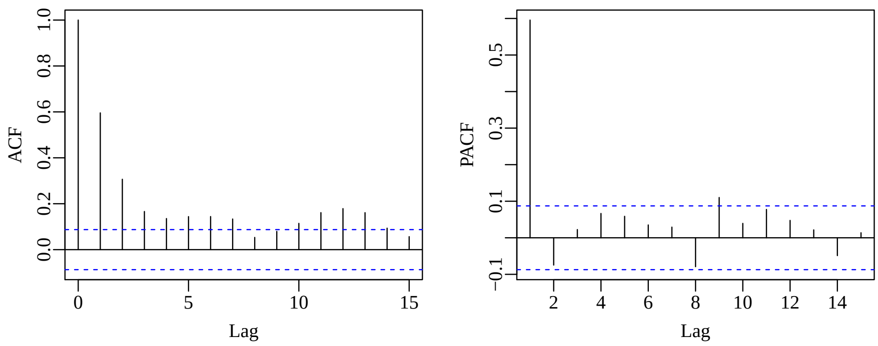


Fig. 7. Number of pedestrians on a city block observed every 5 seconds

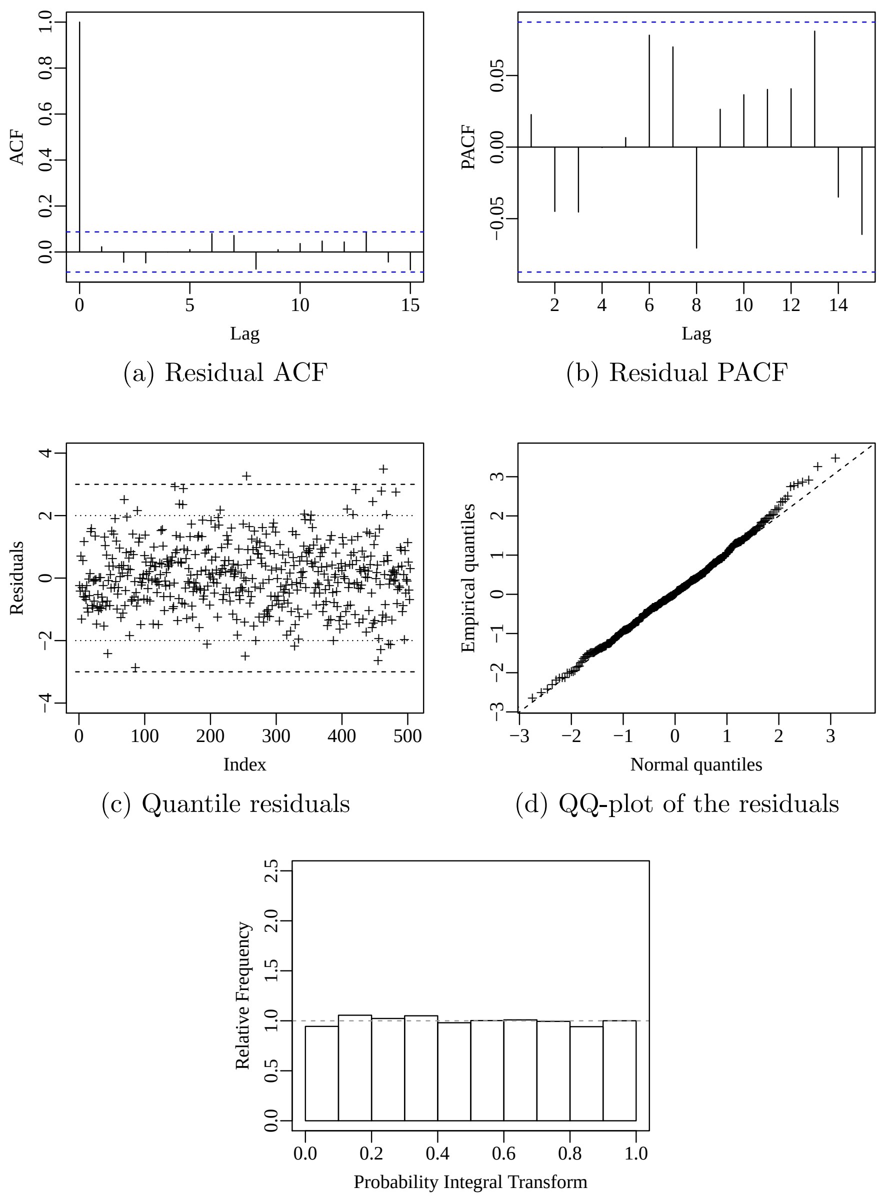


(a) Residual ACF (b) Residual PACF

Fig. 8. (a) ACF and (b) PACF of randomized quantile residuals for the CMP-GLM model.

residuals are normally distributed. Figure 9(e) displays the PIT histogram, indicating that the fitted CMP-ARMA(1,1) model is correctly specified. Figures 10 and 11 present diagnostic plots for the Poisson and Negative GARMA models, respectively. The ACF and PACF of randomized quantile residuals (Figures 10(a), 10(b), 11(a), and 11(b)) indicate that residual autocorrelations are not significant, and Figures 10(c) and 11(c) show that the residuals are randomly distributed around zero for the two models. However, the normal probability and PIT histogram plots (Figures 10(d), 10(e), 11(d), and 11(e)) suggest that the models are not appropriate for these data. All plots indicate overdispersion of the Poisson and Negative Binomial GARMA models. Finally, Figure 12 shows the observed and fitted values for the

CMP-ARMA(1,1) model. The proposed model provides superior fit to the data.



(a) Residual ACF

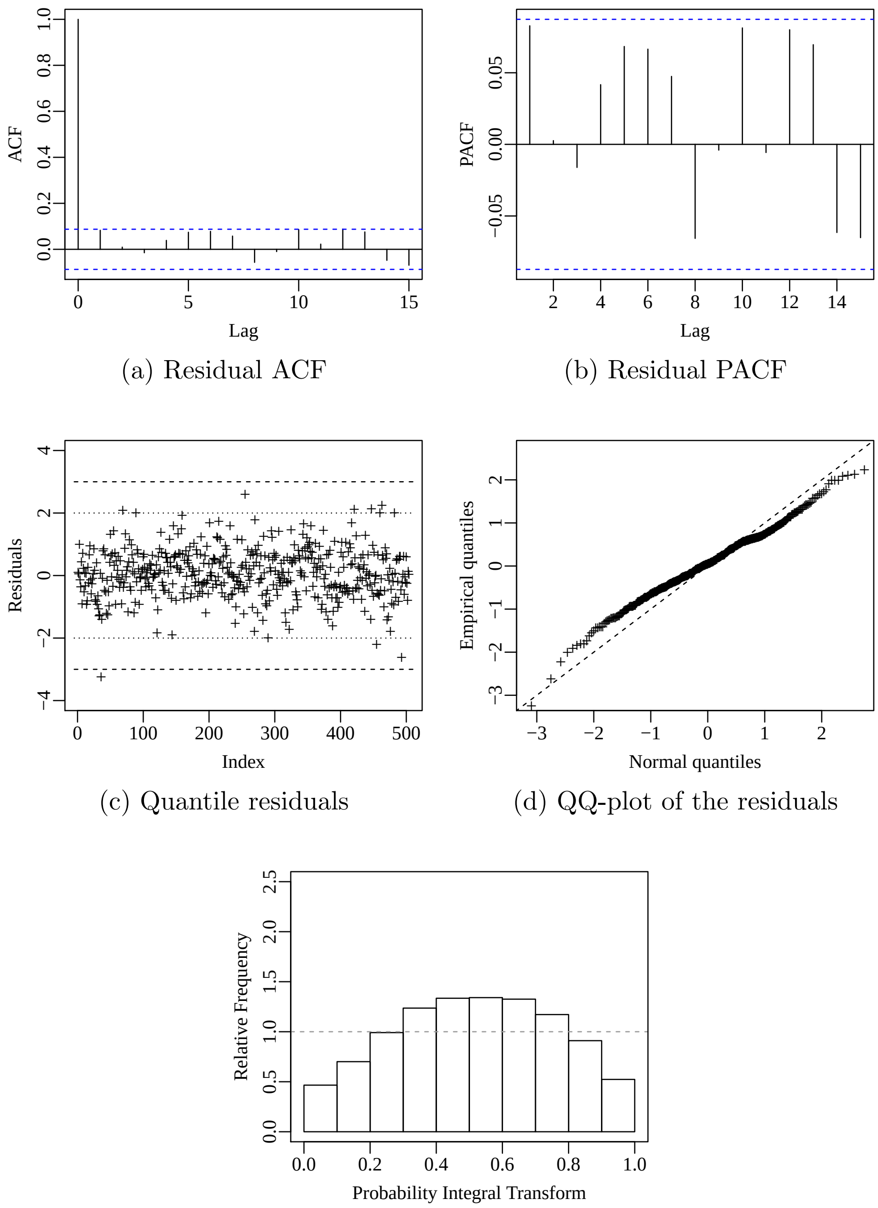
(b) Residual PACF

(c) Quantile residuals

(d) QQ-plot of the residuals

(e) PIT histogram

Fig. 9. Diagnostic plots for the fitted CMP-ARMA model; pedestrians counts data.



(a) Residual ACF (b) Residual PACF

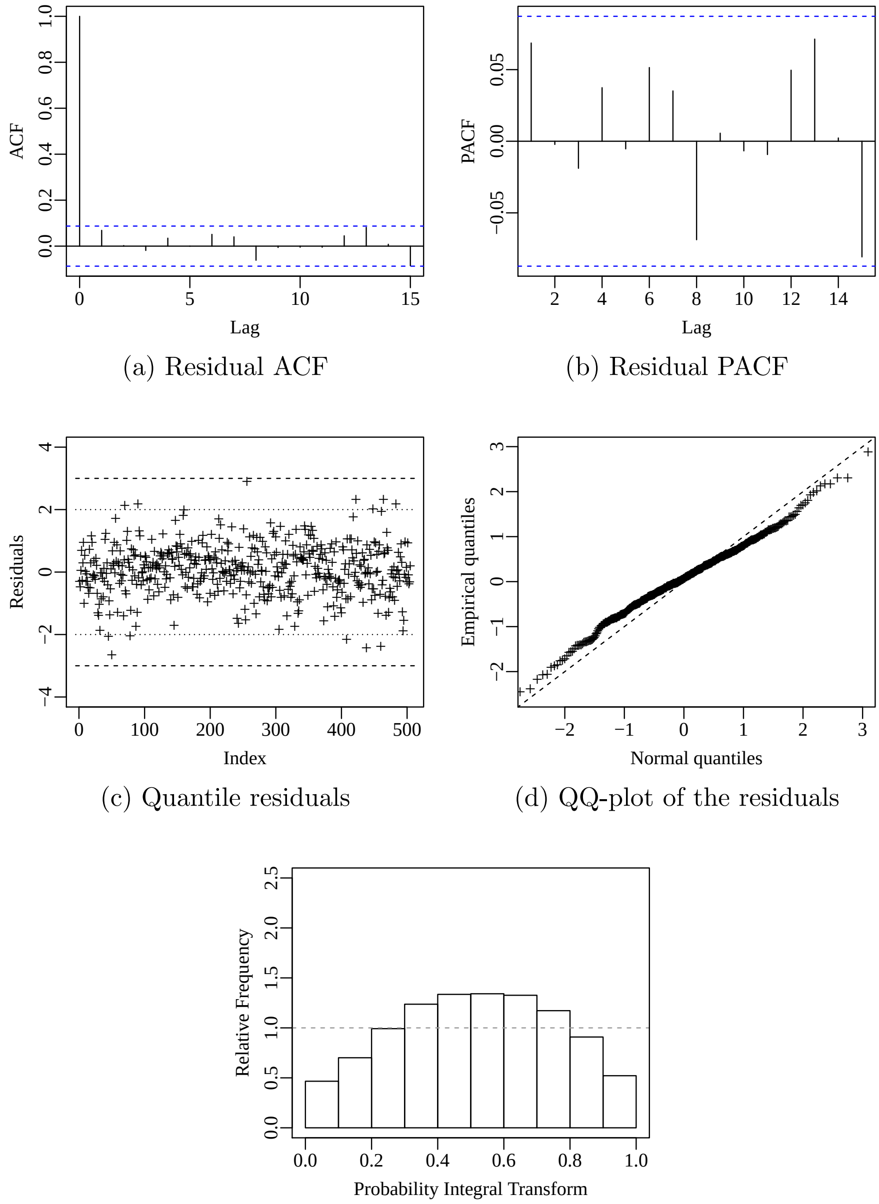
(c) Quantile residuals (d) QQ-plot of the residuals

(e) PIT histogram

Fig. 10. Diagnostic plots for the fitted Poisson GARMA model; pedestrians counts data.

8. Conclusions

The CMP is a flexible distribution that accounts for overdispersion (or underdispersion) encountered in count data. By parametrizing the CMP distribution depending



(a) Residual ACF (b) Residual PACF

(c) Quantile residuals (d) QQ-plot of the residuals

(e) PIT histogram

Fig. 11. Diagnostic plots for the fitted Negative Binomial GARMA model; pedestrians counts data.

on its mean, Huang (2017) proposed a simpler and easily interpretable CMP model, while retaining all the key features of the CMP distributions that have made them increasingly attractive for the analysis of dispersed count data.

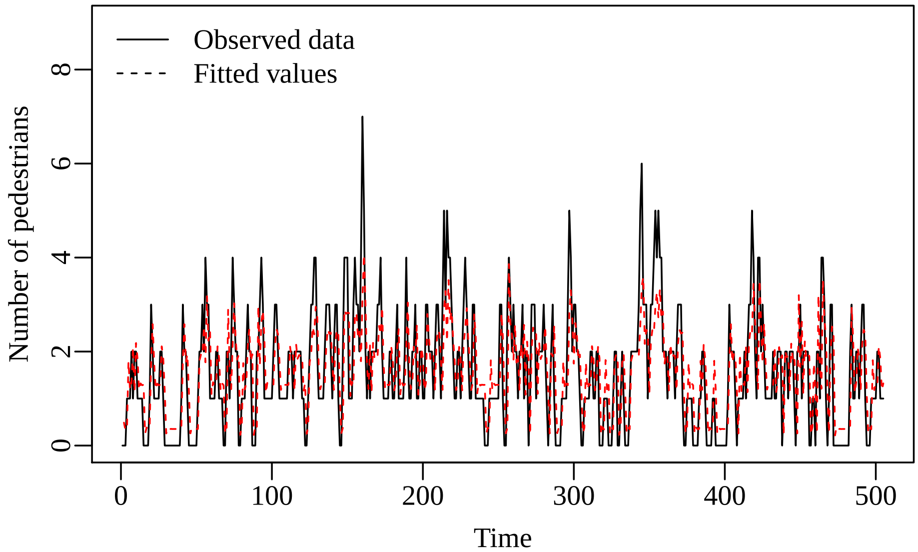


Fig. 12. Fitted and observed values for the CMP-ARMA$(1,1)$ model.

In this paper, we introduced the CMP-ARMA(p,q) dynamic regression model for time series, based on the GARMA model proposed by Benjamin et al. (2003), and we also derived its main properties. The proposed model generalizes the regression model of Huang (2017) by allowing the inclusion of lagged terms to account for autocorrelation. The mean is modeled by a dynamic structure containing autoregressive and moving average terms, time-varying regressors, and a link function. This class of models has potential uses for modeling both underdispersed and overdispersed time series count data.

The model parameters are estimated by the conditional maximum likelihood method. We derived closed-form expressions for the conditional score vector and conditional Fisher information matrix. We also discussed interval estimation, hypothesis testing inference, and model selection criteria. We studied the asymptotic properties of the CMLE in finite samples through a Monte Carlo experiment. We considered the errors measured on the predictor scale $r_t = g(y_t) - g(\mu_t)$. The numerical evidence showed that the CMLE are unbiased and consistent. As in Albarracin et al. (2019), the GARMA model presented an overestimated moving average parameter θ while the autoregressive parameter ϕ was underestimated, indicating that there is multicollinearity between AR and MA terms. It is thus recommended the inclusion of only AR or MA terms to fit the initial model.

Finally, we presented and investigated two empirical applications. In the first data set, with overdispersion, the proposed model performed similarly to the Negative Binomial GARMA model and outperformed the Poisson GARMA model, while for the underdispersed data set (second application) the CMP-ARMA model outperformed both the Poisson and Negative Binomial GARMA models. Our results revealed the importance of the proposed model for count time series since it is capable of modeling overdispersed, equidispersed, and underdispersed data.

Acknowledgments

The authors gratefully acknowledge partial financial support from CAPES, CNPq and FAPESP(2018/04654-9), Brazil. We thank Orlando Yesid Esparza Albarracin for relevant suggestions. We also thank all the corrections and suggestions pointed out by the referees.

Appendix A - Special cases of the CMP-ARMA model

When $\nu = 1$, by solving Equation (4) we have $\lambda_t = \mu_t$ and $Z(\lambda(\mu_t, \nu), \nu) = \sum_{s=0}^{\infty} \lambda(\mu_t, 1)^s/(s!) = e^{-\lambda_t} = e^{-\mu_t}$. Thus, $Pr(Y_t = y_t \mid \mathcal{F}_{t-1}, \mu_t, \nu = 1) = e^{-\mu_t}\mu_t^{y_t}/y_t!$ is the conditional probability of a Poisson distribution with mean μ_t.

When $\nu = 0$, solving Equation (4) gives $\lambda_t = 1/(1+\mu_t) < 1$ and $Z(\lambda(\mu_t, \nu), \nu) = \sum_{s=0}^{\infty} \lambda_t^s = 1/(1-\lambda_t) = (\mu_t + 1)/\mu_t$. Thus, $Pr(Y_t = y_t \mid \mathcal{F}_{t-1}, \mu_t, \nu = 0) = \mu_t/(\mu_t+1)^{y_t+1}$ is the conditional probability of a Geometric distribution with success probability $p_t = 1/(\mu_t + 1)$.

When $\nu \to \infty$, $Z(\lambda(\mu_t, \nu), \nu) \to 1 + \lambda(\mu_t, \nu)$ and the term $(y_t!)^\nu$ in Equation (3) tends to ∞ for $y_t \neq 0, 1$. By solving Equation (4) we have $\lambda_t = \mu_t/1 - \mu_t$. Thus, Y_t, given $\mathcal{F}_{t-1}$, assumes only the values $y_t = 0$ or $y_t = 1$ with probability $1/(1 + \lambda_t) = 1 - \mu_t$ and $\lambda_t/(1 + \lambda_t) = \mu_t$, respectively, and the conditional probability function of Y_t, given $\mathcal{F}_{t-1}$, approaches to a Bernoulli distribution with $Pr(Y_t = 1 \mid \mathcal{F}_{t-1}, \mu_t, \nu \to \infty) = \lambda_t/(1 + \lambda_t) = \mu_t$.

Appendix B - Conditional score vector

We provide below details on the derivation of the following conditional score vector

$$\boldsymbol{U}(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \sum_{t=m}^{n} \frac{\partial \ell_t(\mu_t, \nu)}{\partial \boldsymbol{\gamma}}$$

Part 1 - Score vector for $\boldsymbol{\mu_t}$

By differentiating the conditional log-likelihood function given in (6) with respect to the i-th element of the parameter vector $\boldsymbol{\gamma}$, $\gamma_i \neq \nu$, for $i = 1, \ldots, (r + p + q + 1)$, we obtain

$$U_{\gamma_i}(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_i} = \sum_{t=1}^{n} \frac{\partial \ell_t(\mu_t, \nu)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_i}.$$

Note that $\eta_t = g(\mu_t)$, then $d\mu_t/d\eta_t = 1/g'(\mu_t)$. Next, we shall obtain the derivative of $\ell_t(\mu_t, \nu)$ with respect to μ_t.

First, remember that $Z_t = \sum_{s=0}^{\infty} \lambda_t^s/(s!)^\nu$, where $\lambda_t = \lambda(\mu_t, \nu)$, then

$$\frac{\partial Z_t}{\partial \lambda_t} = \sum_{y_t=0}^{\infty} y_t \frac{\lambda_t^{y_t-1}}{(y_t!)^\nu} = \frac{1}{\lambda_t} \sum_{y_t=0}^{\infty} y_t \frac{\lambda_t^{y_t}}{(y_t!)^\nu} = \frac{1}{\lambda_t} \mu_t Z_t,$$

where μ_t is the mean of the distribution. Therefore, the derivative of the log-likelihood given in the equation (6) with respect to λ_t is given by

$$\frac{\partial \ell_t(\mu_t, \nu)}{\partial \lambda_t} = \frac{y_t}{\lambda_t} - \frac{\partial Z_t/\partial \lambda_t}{Z_t} = \frac{y_t - \mu_t}{\lambda_t}.$$

Second, $\lambda_t(\mu_t, \nu)$ is the solution for $0 = \sum_{y_t=0}^{\infty} (y_t - \mu_t) \lambda_t^{y_t}/(y_t!)^\nu$. Differentiating both sides implicitly with respect to μ, we have

$$\begin{aligned} 0 &= -\left[\sum_{y_t=0}^{\infty} \frac{\lambda_t^{y_t}}{(y_t!)^\nu}\right] + \left[\sum_{y_t=0}^{\infty} (y_t - \mu_t) y_t \frac{\lambda^{y_t-1}}{(y_t!)^\nu}\right] \frac{\partial \lambda_t}{\partial \mu_t} \\ &= -Z_t + \frac{1}{\lambda_t}\left[\sum_{y_t=0}^{\infty} (y_t - \mu_t) y_t \frac{\lambda_t^{y_t}}{(y_t!)^\nu}\right] \frac{\partial \lambda_t}{\partial \mu_t} \\ &= -Z_t + \frac{1}{\lambda_t}\left[\sum_{y_t=0}^{\infty} (y_t - \mu_t)^2 \frac{\lambda^{y_t}}{(y_t!)^\nu}\right] \frac{\partial \lambda_t}{\partial \mu_t}. \end{aligned}$$

The last equality is true because $\sum_{y_t=0}^{\infty} (y_t - \mu_t) \lambda_t^y/(y_t!)^\nu = 0$ by the definition of μ_t as the mean of the distribution. Therefore, we have $\partial \lambda_t/\partial \mu_t = \lambda_t/V(\mu_t, \nu)$, where

$$V(\mu_t, \nu) = \sum_{y_t=0}^{\infty} \frac{(y_t - \mu_t)^2 \lambda_t(\mu_t, \nu)_t^y}{(y_t!)^\nu Z(\lambda_t(\mu_t, \nu), \nu)}$$

is the variance of Y_t.

Finally, we write

$$\frac{\partial \ell_t(\mu_t, \nu)}{\partial \mu_t} = \frac{\partial \ell_t(\mu_t, \nu)}{\partial \lambda(\mu_t, \nu)} \frac{\partial \lambda(\mu_t, \nu)}{\partial \mu_t} = \frac{y_t - \mu_t}{\lambda(\mu, \nu)} \frac{\lambda(\mu, \nu)}{V(\mu, \nu)} = \frac{y_t - \mu_t}{V(\mu_t, \nu)}. \tag{10}$$

Therefore, it follows that

$$U_{\gamma_i}(\boldsymbol{\gamma}) = \sum_{t=1}^{n} \frac{y_t - \mu_t}{V(\mu_t, \nu) g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_i}, \text{ for } \gamma_i \notin \nu. \tag{11}$$

Define the error in (5) by $r_t = g(y_t) - g(\mu_t)$. When computing the derivative of

η_t with respect to $\gamma_i \neq \nu$, we obtain

$$\begin{aligned}
\frac{\partial \eta_t}{\partial \alpha} =& 1 + \sum_{j=1}^{q} \theta_j \frac{\partial r_{t-j}}{\partial \alpha} = 1 - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \alpha}, \\
\frac{\partial \eta_t}{\partial \beta_i} =& x_{ti} - \sum_{j=1}^{p} \phi_j x_{(t-j)i} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \beta_i}, \\
\frac{\partial \eta_t}{\partial \phi_i} =& g(y_{t-i}) - \boldsymbol{x}_{\boldsymbol{t-i}}^{\top} \boldsymbol{\beta} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \phi_i}, \\
\frac{\partial \eta_t}{\partial \theta_i} =& (g(y_{t-i}) - \eta_{t-i}) - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \theta_i}.
\end{aligned}$$

Part 2 - Score vector for ν

Since $\lambda_t(\mu_t, \nu)$ is the solution for $0 = \sum_{y_t=0}^{\infty} (y_t - \mu_t) \lambda_t^y / (y_t!)^\nu$, by differentiating both sides implicitly with respect to ν we have

$$\begin{aligned}
0 &= -\left[\sum_{y_t=0}^{\infty} (y_t - \mu_t) \frac{\log(y_t!) \lambda_t^{y_t}}{(y_t!)^\nu} \right] + \left[\sum_{y_t=0}^{\infty} (y_t - \mu_t) y_t \frac{\lambda_t^{y_t - 1}}{(y_t!)^\nu} \right] \frac{\partial \lambda_t}{\partial \nu} \\
&= -\left[\sum_{y_t=0}^{\infty} (y_t - \mu_t) \frac{\log(y_t!) \lambda_t^{y_t}}{(y_t!)^\nu} \right] + \frac{1}{\lambda_t} \left[\sum_{y_t=0}^{\infty} (y_t - \mu_t) y_t \frac{\lambda_t^{y_t}}{(y_t!)^\nu} \right] \frac{\partial \lambda_t}{\partial \nu} \\
&= -\left[\sum_{y_t=0}^{\infty} (y_t - \mu_t) \frac{\log(y_t!) \lambda_t^{y_t}}{(y_t!)^\nu Z_t} \right] + \frac{1}{\lambda_t} \left[\sum_{y_t=0}^{\infty} (y_t - \mu_t)^2 \frac{\lambda_t^{y_t}}{(y_t!)^\nu Z_t} \right] \frac{\partial \lambda_t}{\partial \nu} \\
&= -E_{\mu_t, \nu} \left[\log(y_t!)(y_t - \mu_t) \right] + \frac{1}{\lambda_t} \left[V(\mu_t, \nu) \right] \frac{\partial \lambda_t}{\partial \nu} \\
\frac{\partial \lambda_t}{\partial \nu} &= \frac{\lambda_t E_{\mu_t, \nu} \left[\log(y_t!)(y_t - \mu_t) \right]}{V(\mu_t, \nu)}.
\end{aligned}$$

Remember also that $Z_t = \sum_{y_t=0}^{\infty} \lambda_t^{y_t} / (y_t!)^\nu$, then

$$\begin{aligned}
\frac{\partial Z_t}{\partial \nu} &= \sum_{y_t=0}^{\infty} \left[y_t \frac{\lambda_t^{y_t - 1}}{(y_t!)^\nu} \frac{\partial \lambda_t}{\partial \nu} - \frac{\lambda_t^{y_t} \log(y_t!)}{(y_t!)^\nu} \right] \\
&= \frac{E_{\mu_t, \nu} \left[\log(y_t!)(y_t - \mu_t) \right]}{V(\mu_t, \nu)} \sum_{y_t=0}^{\infty} y_t \frac{\lambda_t^{y_t}}{(y_t!)^\nu} - \sum_{y_t=0}^{\infty} \frac{\lambda_t^{y_t} \log(y_t!)}{(y_t!)^\nu}.
\end{aligned}$$

Thus, the derivative of the log-likelihood given in Equation (6) with respect to ν is given by

$$\begin{aligned}\frac{\partial \ell_t(\mu_t,\nu)}{\partial \nu} &= y_t \frac{\partial \lambda_t/\partial \nu}{\lambda_t} - \log(y_t) - \frac{\partial Z_t/\partial \nu}{Z_t}\\ &= y_t \frac{E_{\mu_t,\nu}\left[\log(y_t!)(y_t-\mu_t)\right]}{V(\mu_t,\nu)} - \log(y_t) - \left[\mu_t \frac{E_{\mu_t,\nu}\left[\log(y_t!)(y_t-\mu_t)\right]}{V(\mu_t,\nu)}\right.\\ &\qquad\qquad \left. - E_{\mu_t,\nu}\log(y_t!)\right]\\ &= E_{\mu_t,\nu}\left[\log(y_t!)(y_t-\mu_t)\right]\frac{(y_t-\mu_t)}{V(\mu_t,\nu)} - \left[\log(y_t!) - E_{\mu_t,\nu}\log(y_t!)\right].\end{aligned}$$

Therefore, we obtain

$$\frac{\partial \ell_t(\mu_t,\nu)}{\partial \nu} = \sum_{t=1}^{n} A(\mu_t,\nu)\frac{(y_t-\mu_t)}{V(\mu_t,\nu)} - [\log(y_t!) - B(\mu_t,\nu)], \qquad (12)$$

where $A(\mu_t,\nu) = E_{\mu_t,\nu}\left[\log(y_t!)(y_t-\mu_t)\right]$ and $B(\mu_t,\nu) = E_{\mu_t,\nu}\log(y_t!)$.

From (11) and (12), we then obtain the matrix expression for the score vector given in (7).

Appendix C - Conditional information matrix

In this appendix we derive the conditional Fisher information matrix for $\boldsymbol{\gamma}$. In order to make algebra easier, we initially present some preliminary results.

Part 1 - Derivative of $V(\mu_t,\nu)$ with respect to μ_t

Firstly, $Z_t = \sum_{y_t=0}^{\infty} \lambda_t^y/(y_t!)^\nu$, then

$$\frac{\partial Z_t}{\partial \mu_t} = \sum_{y_t=0}^{\infty} y_t \frac{\lambda_t^{y_t-1}}{(y_t!)^\nu}\frac{\partial \lambda_t}{\partial \mu_t} = \frac{1}{V(\mu_t,\nu)}\sum_{y_t=0}^{\infty} y_t \frac{\lambda^{y_t}}{(y_t!)^\nu} = \frac{\mu_t Z_t}{V(\mu_t,\nu)}.$$

Let $V(\mu_t,\nu) = \sum_{y_t=0}^{\infty} \frac{(y_t-\mu_t)^2\lambda^{y_t}}{(y_t!)^\nu Z_t}$, then the derivative of $V(\mu_t,\nu)$ with respect to μ_t is given by

$$\begin{aligned}\frac{\partial V(\mu_t,\nu)}{\partial \mu_t} &= \sum_{y_t=0}^{\infty}\left[\frac{-2\mu_t(y_t-\mu_t)\lambda_t^{y_t}(y_t!)^\nu Z_t}{(y_t!)^{2\nu}Z_t^2} + \frac{(y_t-\mu_t)^2 y_t \lambda_t^{y_t-1}(\partial\lambda_t/\partial\mu_t)(y_t!)^\nu Z_t}{(y_t!)^{2\nu}Z_t^2}\right.\\ &\qquad \left. - \frac{(y_t-\mu_t)^2\lambda_t^{y_t}(y_t!)^\nu(\partial Z_t/\partial\mu_t)}{(y_t!)^{2\nu}Z_t^2}\right]\\ &= \sum_{y_t=0}^{\infty}\left[\frac{-2\mu_t(y_t-\mu_t)\lambda_t^{y_t}}{(y_t!)^\nu Z_t} + \frac{(y_t-\mu_t)^2 y_t\lambda_t^{y_t}}{V(\mu_t,\nu)(y_t!)^\nu Z_t} - \frac{(y_t-\mu_t)^2\mu_t\lambda_t^{y_t}}{V(\mu_t,\nu)(y_t!)^\nu Z_t}\right]\\ &= \frac{1}{V(\mu_t,\nu)}\sum_{y_t=0}^{\infty}\left[\frac{(y_t-\mu_t)^3\lambda_t^{y_t}}{(y_t!)^\nu Z_t}\right] = \frac{m_3(\mu_t,\nu)}{V(\mu_t,\nu)},\end{aligned}$$

where $m_3(\mu_t, \nu)$ is the third central moment.

Part 2 - Derivative of $B(\mu_t,\nu)$ with respect to μ_t

Secondly, let $B(\mu_t, \nu) = \sum_{y_t=0}^{\infty} \frac{\log(y_t!)\lambda_t^{y_t}}{(y_t!)^\nu Z_t}$. The derivative of $B(\mu_t, \nu)$ with respect to μ_t is defined as

$$\begin{aligned}
\frac{\partial B(\mu_t, \nu)}{\partial \mu_t} &= \sum_{y_t=0}^{\infty} \left[\frac{\log(y_t!) y_t \lambda_t^{y_t - 1} (\partial \lambda_t / \partial \mu_t)(y_t!)^\nu Z_t}{(y_t!)^{2\nu} Z_t^2} - \frac{\log(y_t!) \lambda_t^{y_t} (y_t!)^\nu (\partial Z_t / \partial \mu_t)}{(y_t!)^{2\nu} Z_t^2} \right] \\
&= \sum_{y_t=0}^{\infty} \left[\frac{\log(y_t!) y_t \lambda^{y_t}}{V(\mu_t, \nu)(y_t!)^\nu Z_t} - \frac{\mu_t \log(y_t!) \lambda_t^{y_t}}{V(\mu_t, \nu)(y_t!)^\nu Z_t} \right] \\
&= \frac{1}{V(\mu_t, \nu)} \sum_{y_t=0}^{\infty} \frac{\log(y_t!)(y_t - \mu_t) \lambda_t^{y_t}}{(y_t!)^\nu Z_t} \\
&= \frac{A(\mu_t, \nu)}{V(\mu_t, \nu)}.
\end{aligned}$$

Part 3 - Derivative of $A(\mu_t,\nu)$ with respect to μ_t

Thirdly, let $A(\mu_t, \nu) = \sum_{y_t=0}^{\infty} \frac{\log(y_t!)(y_t - \mu_t)\lambda_t^{y_t}}{(y_t!)^\nu Z_t}$, then the derivative of $A(\mu_t, \nu)$ with respect to μ_t is

$$\begin{aligned}
\frac{\partial A(\mu_t, \nu)}{\partial \mu_t} &= \sum_{y_t=0}^{\infty} \left[\frac{[\log(y_t!) \lambda^{y_t} + \log(y_t!)(y_t - \mu_t) y_t \lambda_t^{y_t - 1} (\partial \lambda_t / \partial \mu_t)](y_t!)^\nu Z_t}{(y_t!)^{2\nu} Z_t^2} \right. \\
&\qquad \left. - \frac{(y_t!)^\nu (\partial Z_t / \partial \mu_t) \log(y_t!)(y_t - \mu_t) \lambda_t^{y_t}}{(y_t!)^{2\nu} Z_t^2} \right] \\
&= \sum_{y_t=0}^{\infty} \left[\frac{\log(y_t!) \lambda_t^{y_t}}{(y_t!)^\nu Z_t} + \frac{\log(y_t!)(y_t - \mu_t) y_t \lambda_t^{y_t}}{V(\mu_t, \nu)(y_t!)^\nu Z_t} - \frac{\mu_t \log(y_t!)(y_t - \mu_t) \lambda_t^{y_t}}{V(\mu_t, \nu)(y_t!)^\nu Z_t} \right] \\
&= \sum_{y_t=0}^{\infty} \left[\frac{\log(y_t!) \lambda_t^{y_t}}{(y_t!)^\nu Z_t} + \frac{\log(y_t!)(y_t - \mu_t)^2 \lambda_t^{y_t}}{V(\mu_t, \nu)(y_t!)^\nu Z_t} \right] \\
&= E_{\mu_t,\nu} \log(y_t!) + \frac{E_{\mu_t,\nu}[\log(y_t!)(y_t - \mu_t)^2]}{V(\mu_t, \nu)} \\
&= B(\mu_t, \nu) + \frac{D(\mu_t, \nu)}{V(\mu_t, \nu)},
\end{aligned}$$

where $D(\mu_t, \nu) = E_{\mu_t,\nu}[\log(y!)(y - \mu)^2]$.

Part 4 - Derivative of $A(\mu_t,\nu)/V(\mu_t,\nu)$ with respect to μ_t

Finally, by deriving $A(\mu_t,\nu)/V(\mu_t,\nu)$ with respect to μ_t, the quotient rule yields

$$\begin{aligned}\frac{\partial}{\partial\mu_t}\left[\frac{A(\mu_t,\nu)}{V(\mu_t,\nu)}\right] &= \frac{[\partial A(\mu_t,\nu)/\partial\mu_t)V(\mu_t,\nu]}{V(\mu_t,\nu)^2} - \frac{[\partial V(\mu_t,\nu)/\partial\mu_t]A(\mu_t,\nu)}{V(\mu_t,\nu)^2} \\ &= \frac{B(\mu_t,\nu)}{V(\mu_t,\nu)} + \frac{D(\mu_t,\nu)}{V(\mu_t,\nu)^2} - \frac{m_3(\mu_t,\nu)}{V(\mu_t,\nu)^3} = F(\mu_t,\nu).\end{aligned}$$

Part 5 - Derivatives of $E_{\mu_t,\nu}\log(y_t!)$ with respect to ν

Let $E_{\mu_t,\nu}\log(y_t!) = \sum_{y_t=0}^{\infty}\frac{\log(y_t!)\lambda^{y_t}}{(y_t!)^\nu Z_t}$, then the derivative of $E_{\mu_t,\nu}\log(y_t!)$ with respect to ν is given by

$$\begin{aligned}\frac{\partial E_{\mu_t,\nu}\log(y_t!)}{\partial\nu} &= \sum_{y_t=0}^{\infty}\left[\frac{\log(y_t!)y\lambda^{y_t-1}(\partial\lambda_t/\partial\nu)(y_t!)^\nu Z_t}{(y_t!)^{2\nu}Z_t^2} - \frac{\log(y_t!)\lambda^{y_t}(y_t!)^\nu\log(y_t!)Z_t}{(y_t!)^{2\nu}Z_t^2}\right. \\ &\qquad\left. - \frac{\log(y_t!)\lambda_t^{y_t}(y_t!)^\nu(\partial Z_t/\partial\nu)}{(y_t!)^{2\nu}Z_t^2}\right] \\ &= \sum_{y_t=0}^{\infty}\left[\frac{\log(y_t!)y_t\lambda_t^{y_t}E_{\mu_t,\nu}\left[\log(y_t!)(y_t-\mu_t)\right]}{V(\mu_t,\nu)(y_t!)^\nu Z_t} - \frac{\log(y_t!)^2\lambda_t^{y_t}}{(y_t!)^\nu Z_t}\right. \\ &\qquad\left. - \frac{\log(y_t!)\lambda_t^{y_t}}{(y_t!)^\nu Z_t}\left(\mu_t\frac{E_{\mu_t,\nu}\left[\log(y_t!)(y_t-\mu_t)\right]}{V(\mu_t,\nu)} - E_{\mu_t,\nu}\log(y_t!)\right)\right] \\ &= \frac{E_{\mu_t,\nu}\left[\log(y_t!)(y_t-\mu_t)\right]}{V(\mu_t,\nu)}\sum_{y_t=0}^{\infty}\frac{(y_t-\mu_t)\log(y_t!)\lambda_t^{y_t}}{(y_t!)^\nu Z_t} - \sum_{y_t=0}^{\infty}\frac{\log(y_t!)^2\lambda_t^{y_t}}{(y_t!)^\nu Z_t} \\ &\qquad + E_{\mu_t,\nu}\log(y_t!)\sum_{y_t=0}^{\infty}\frac{\log(y_t!)\lambda_t^{y_t}}{(y_t!)^\nu Z_t} \\ &= \frac{E_{\mu_t,\nu}\left[\log(y_t!)(y_t-\mu_t)\right]^2}{V(\mu_t,\nu)} - \left[E_{\mu_t,\nu}\log(y_t!)^2 - (E_{\mu_t,\nu}\log(y_t!))^2\right] \\ &= \frac{A(\mu_t,\nu)^2}{V(\mu_t,\nu)} - C(\mu_t,\nu),\end{aligned}$$

where $C(\mu_t,\nu) = V_{\mu_t,\nu}(\log(y_t))$.

We shall now derive the conditional Fisher information matrix for the proposed. We have

$$K(\boldsymbol{\gamma}) = E\left[-\frac{\partial^2\ell(\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right] = E\left[-\sum_{t=m}^{n}\frac{\partial^2\ell_t(\mu_t,\nu)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right].$$

For $\gamma_i \neq \nu$ and $\gamma_j \neq \nu$, for $i,j \in \{1,\ldots,r+p+q+1\}$, we obtain

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j} &= \sum_{t=m+1}^{n} \frac{\partial}{\partial \gamma_i}\left[\frac{\partial \ell_t(\mu_t,\nu)}{\partial \mu_t}\frac{d\mu_t}{d\eta_t}\frac{\partial \eta_t}{\partial \gamma_j}\right] \\
&= \sum_{t=m+1}^{n}\left[\frac{\partial^2 \ell_t(\mu_t,\nu)}{\partial \mu_t^2}\frac{d\mu_t}{d\eta_t}\frac{\partial \eta_t}{\partial \gamma_j}\frac{d\mu_t}{d\eta_t}\frac{\partial \eta_t}{\partial \gamma_i} + \frac{\partial \ell_t(\mu_t,\nu)}{\partial \mu_t}\frac{\partial}{\partial \gamma_i}\left(\frac{d\mu_t}{d\eta_t}\frac{\partial \eta_t}{\partial \gamma_j}\right)\right] \\
&= \sum_{t=m+1}^{n}\left[\frac{\partial^2 \ell_t(\mu_t,\nu)}{\partial \mu_t^2}\frac{d\mu_t}{d\eta_t}\frac{\partial \eta_t}{\partial \gamma_j}\frac{d\mu_t}{d\eta_t}\frac{\partial \eta_t}{\partial \gamma_i} + \frac{\partial \ell_t(\mu_t,\nu)}{\partial \mu_t}\frac{\partial \eta_t}{\partial \gamma_j}\frac{d^2\mu_t}{d\eta_t^2}\frac{\partial \eta_t}{\partial \gamma_i}\right. \\
&\qquad \left. + \frac{\partial \ell_t(\mu_t,\nu)}{\partial \mu_t}\frac{d\mu_t}{d\eta_t}\frac{\partial^2 \eta_t}{\partial \gamma_j \partial \gamma_i}\right].
\end{aligned}$$

Let $\partial \ell_t(\mu_t,\nu)/\partial \mu_t | \mathcal{F}_{t-1}$ given by (10), so it follows that $E\left[\frac{\partial \ell_t(\mu_t,\nu)}{\partial \mu_t}\Big|\mathcal{F}_{t-1}\right] = 0$. Therefore,

$$E\left[\frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}\Big|\mathcal{F}_{t-1}\right] = \sum_{t=1}^{n} E\left[\frac{\partial^2 \ell_t(\mu_t,\nu)}{\partial \mu_t^2}\Big|\mathcal{F}_{t-1}\right]\left(\frac{d\mu_t}{d\eta_t}\right)^2 \frac{\partial \eta_t}{\partial \gamma_i}\frac{\partial \eta_t}{\partial \gamma_j}. \qquad (13)$$

From (10) we also obtain

$$\begin{aligned}
\frac{\partial^2 \ell_t(\mu_t,\nu)}{\partial \mu_t^2} &= \frac{\partial}{\partial \mu}\left[\frac{y_t-\mu_t}{V(\mu_t,\nu)}\right] = \frac{-V(\mu_t,\nu) - (y_t-\mu_t)\left[\partial V(\mu_t,\nu)/\partial \mu_t\right]}{V(\mu_t,\nu)^2} \\
&= \frac{-V(\mu_t,\nu) - (y_t-\mu_t)\left[m_3(\mu_t,\nu)/V(\mu_t,\nu)\right]}{V(\mu_t,\nu)^2}.
\end{aligned}$$

Thus,

$$E\left[\frac{\partial^2 \ell_t(\mu_t,\nu)}{\partial \mu_t^2}\Big|\mathcal{F}_{t-1}\right] = \frac{-1}{V(\mu_t,\nu)}. \qquad (14)$$

By replacing (14) in (13) it follows that

$$E\left[\frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}\Big|\mathcal{F}_{t-1}\right] = \sum_{t=1}^{n} \frac{-1}{V(\mu_t,\nu) g'(\mu_t)^2}\frac{\partial \eta_t}{\partial \gamma_i}\frac{\partial \eta_t}{\partial \gamma_j}. \qquad (15)$$

From (15) we obtain the information matrix using the derivatives of η_t with respect to the parameters ($\gamma_i \neq \nu$ e $\gamma_j \neq \nu$) previously obtained in 4.1.

Now, derivatives with respect to ν, are easily obtained directly as follows

$$\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \nu} = \sum_{t=1}^{n} A(\mu_t,\nu)\frac{(y_t-\mu_t)}{V(\mu_t,\nu)} - \left[\log(y_t!) - B(\mu_t,\nu)\right],$$

then,

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \nu \partial \gamma_j} &= \sum_{t=1}^{n} \left[\frac{\partial [A(\mu_t,\nu)/V(\mu_t,\nu)]}{\partial \gamma_j}(y_t-\mu_t) - \frac{A(\mu_t,\nu)}{V(\mu_t,\nu)}\frac{\partial \mu_t}{\partial \gamma_j} + \frac{\partial B(\mu_t,\nu)}{\partial \gamma_j} \right] \\
&= \sum_{t=1}^{n} \left[\frac{\partial [A(\mu_t,\nu)/V(\mu_t,\nu)]}{\partial \gamma_j}(y_t-\mu_t) - \frac{A(\mu_t,\nu)}{V(\mu_t,\nu)}\frac{\partial \mu_t}{\partial \gamma_j} + \frac{\partial B(\mu_t,\nu)}{\partial \mu_t}\frac{\partial \mu_t}{\partial \gamma_j} \right] \\
&= \sum_{t=1}^{n} \left[\frac{\partial [A(\mu_t,\nu)/V(\mu_t,\nu)]}{\partial \gamma_j}(y_t-\mu_t) - \frac{A(\mu_t,\nu)}{V(\mu_t,\nu)}\frac{\partial \mu_t}{\partial \gamma_j} + \frac{A(\mu_t,\nu)}{V(\mu_t,\nu)}\frac{\partial \mu_t}{\partial \gamma_j} \right] \\
&= \sum_{t=1}^{n} \left[\frac{\partial [A(\mu_t,\nu)/V(\mu_t,\nu)]}{\partial \mu_t}\frac{\partial \mu_t}{\partial \gamma_j}(y_t-\mu_t) \right] \\
&= \sum_{t=1}^{n} \left[F(\mu_t,\nu)\frac{\partial \mu_t}{\partial \gamma_j}(y_t-\mu_t) \right].
\end{aligned}
$$

Since $E(Y_t|\mathcal{F}_{t-1}) = \mu_t$, we have

$$E\left[\frac{\partial^2 \ell(\mu_t,\nu)}{\partial \nu \partial \gamma_i}\middle|\mathcal{F}_{t-1}\right] = 0,$$

that is, ν is orthogonal to the other parameters. Finally, we obtain

$$\frac{\partial \ell}{\partial \nu} = \boldsymbol{U}_\nu(\boldsymbol{\gamma}) = \sum_{t=1}^{n} E_{\mu_t,\nu}\left[\log(y_t!)(\mu_t - y)\right]\frac{(y_t-\mu_t)}{V(\mu_t,\nu)} - \left[\log(y_t!) - E_{\mu_t,\nu}\log(y_t!)\right].$$

As we are interested in $E\left(\frac{\partial^2 \ell}{\partial \nu^2}\middle|\mathcal{F}_{t-1}\right)$, it follows that by deriving $\boldsymbol{U}_\nu(\boldsymbol{\gamma})$ and then applying the expectation, the first and second terms will be zero. Thus,

$$E\left[\frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \nu^2}\middle|\mathcal{F}_{t-1}\right] = \sum_{t=1}^{n} \frac{\partial E_{\mu,\nu}\log(y!)}{\partial \nu} = \sum_{t=1}^{n}\left[\frac{A(\mu_t,\nu)^2}{V(\mu_t,\nu)} - C(\mu_t,\nu)\right].$$

Obtaining the Fisher information matrix for $\boldsymbol{\gamma}$ is now an easy task.

Data availability statement

The empirical data sets used in Section 7 are available in the Supporting Information

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article.

References

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on 19*(6), 716–723.

Albarracin, O. Y. E., A. P. Alencar, and L. L. Ho (2019). Generalized autoregressive and moving average models: multicollinearity, interpretation and a new modified model. *Journal of Statistical Computation and Simulation 89*(10), 1819–1840.

Alencar, A. P. (2018). Seasonality of hospitalizations due to respiratory diseases: modelling serial correlation all we need is Poisson. *Journal of Applied Statistics 45*(10), 1813–1822.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society. Series B (Methodological)*, 283–301.

Bayer, F. M., D. M. Bayer, and G. Pumi (2017). Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology 555*, 385–396.

Benjamin, M. A., R. A. Rigby, and D. M. Stasinopoulos (2003). Generalized autoregressive moving average models. *Journal of the American Statistical association 98*(461), 214–223.

Benjamin, M. A., R. A. Rigby, and M. D. Stasinopoulos (1998). Fitting non-gaussian time series models. In *COMPSTAT Proceedings in Computational Statistics Heidelburg: Physica-Verlag*, pp. 191–196.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Conway, R. W. and W. L. Maxwell (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering 12*(2), 132–136.

Czado, C., T. Gneiting, and L. Held (2009). Predictive model assessment for count data. *Biometrics 65*(4), 1254–1261.

Dunn, P. K. and G. K. Smyth (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics 5*(3), 236–244.

Evans, D. (1953). Experimental evidence concerning contagious distributions in ecology. *Biometrika 40*(1/2), 186–211.

Feng, C., A. Sadeghpour, and L. Li (2017). Randomized quantile residuals: an omnibus model diagnostic tool with unified reference distribution. *arXiv preprint arXiv:1708.08527*.

Fokianos, K. and B. Kedem (2004). Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis 25*(2), 173–197.

Franke, J. and T. Seligmann (1993). Conditional maximum likelihood estimates for INAR (1) processes and their application to modelling epileptic seizure counts. *Developments in time series 310*, 330.

Freeland, R. and B. P. McCabe (2004). Analysis of low count time series data by poisson autoregression. *Journal of Time Series Analysis 25*(5), 701–722.

Fürth, R. (1918). Statistik und wahrscheinlichkeitsnachwirkung. *Physikalische Zeitschrift 19*, 421–426.

Gay, D. M. (1990). Usage summary for selected optimization routines. *Computing science technical report 153*, 1–21.

Huang, A. (2017). Mean-parametrized Conway–Maxwell–Poisson regression models for dispersed counts. *Statistical Modelling 17*(6), 359–380.

Jung, R., B. McCabe, and A. Tremayne (2015). Model validation and diagnostics. in handbook of discrete-valued time series.

Jung, R. C. and A. Tremayne (2006). Binomial thinning models for integer time series. *Statistical Modelling 6*(2), 81–96.

Jung, R. C. and A. Tremayne (2011). Useful models for time series of counts or simply wrong ones? *AStA Advances in Statistical Analysis 95*(1), 59–91.

Kedem, B. and K. Fokianos (2005). *Regression models for time series analysis*, Volume 488. John Wiley & Sons.

Li, W. (1991). Testing model adequacy for some markov regression models for time series. *Biometrika 78*(1), 83–89.

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

Liboschik, T., K. Fokianos, and R. Fried (2017). tscount: An r package for analysis of count time series following generalized linear models. *Journal of Statistical Software 82*(5), 1–51.

Ljung, G. M. and G. E. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

MacDonald, I. L. and F. Bhamani (2018). A time-series model for underdispersed or overdispersed counts. *The American Statistician*, 1–12.

Mamode Khan, N., V. Jowaheer, Y. Sunecher, and M. Bourguignon (2018). Modeling longitudinal INMA (1) with COM–Poisson innovation under non-stationarity: application to medical data. *Computational and Applied Mathematics*, 1–22.

McCullagh, P. and J. Nelder (1989). *Generalized linear models* (2nd ed.). Chapman and Hall.

Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rocha, A. V. and F. Cribari-Neto (2009). Beta autoregressive moving average models. *Test 18*(3), 529–545.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics 6*(2), 461–464.

Sellers, K. F., S. Borle, and G. Shmueli (2012). The COM-Poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry 28*(2), 104–116.

Sellers, K. F. and G. Shmueli (2010). A flexible regression model for count data. *The Annals of Applied Statistics*, 943–961.

Shmueli, G., T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright (2005). A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 54*(1), 127–142.

Stasinopoulos, M. M. and B. Rigby (2016). Package gamlss.util. *page 9*.

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

Zhu, F. (2012). Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling 56*(9-10), 191–203.

Table 1. Monte Carlo simulation results for CMLE on CMP-ARMA(1,1) model.

Scenario 1 - overdispersion						
		α	β_1	ϕ_1	θ_1	ν
	Parameters	1.5	0.5	0.5	0.3	0.5
n = 50	Mean	1.6873	0.5052	0.4350	0.3189	0.5757
	RB(%)	12.4923	1.0419	−12.9926	6.3083	15.1501
	MSE	0.3605	0.0129	0.0404	0.0489	0.0463
n = 100	Mean	1.5997	0.5012	0.4655	0.3099	0.5350
	RB(%)	6.6471	0.2445	−6.8854	3.3208	7.0074
	MSE	0.1457	0.0062	0.0162	0.0190	0.0156
n = 200	Mean	1.5518	0.5003	0.4822	0.3056	0.5169
	RB(%)	3.4570	0.0491	−3.5533	1.8748	3.3947
	MSE	0.0634	0.0031	0.0070	0.0084	0.0062
n = 400	Mean	1.5269	0.5003	0.4908	0.3025	0.5082
	RB(%)	1.7935	0.0491	−1.8387	0.8569	1.6499
	MSE	0.0313	0.0015	0.0034	0.0039	0.0028
Scenario 2 - equidispersion						
		α	β_1	ϕ_1	θ_1	ν
	Parameters	1.5	0.5	0.5	0.3	1.0
n = 50	Mean	1.6937	0.5040	0.4336	0.3208	1.1494
	RB(%)	12.9184	0.8036	−13.2679	6.9661	14.9459
	MSE	0.3720	0.0090	0.0416	0.0507	0.0918
n = 100	Mean	1.6029	0.5009	0.4648	0.3111	1.0689
	RB(%)	6.8606	0.1979	−7.0274	3.6943	6.8913
	MSE	0.1535	0.0043	0.0170	0.0200	0.0314
n = 200	Mean	1.5520	0.5009	0.4823	0.3059	1.0326
	RB(%)	3.4692	0.1979	−3.5334	1.9905	3.2641
	MSE	0.0664	0.0021	0.0073	0.0089	0.0125
n = 400	Mean	1.5266	0.5007	0.4909	0.3031	1.0157
	RB(%)	1.7762	0.1584	−1.8077	1.0493	1.5766
	MSE	0.0331	0.0011	0.0036	0.0041	0.0058
Scenario 3 - underdispersion						
		α	β_1	ϕ_1	θ_1	ν
	Parameters	1.5	0.5	0.5	0.3	2.0
n = 50	Mean	1.7001	0.5021	0.4324	0.3214	2.2982
	RB(%)	13.3400	0.4378	−13.5150	7.1332	14.9109
	MSE	0.3845	0.0049	0.0428	0.0536	0.3661
n = 100	Mean	1.6079	0.5006	0.4636	0.3119	2.1367
	RB(%)	7.1985	0.1298	−7.2750	3.9884	6.8370
	MSE	0.1617	0.0021	0.0179	0.0209	0.1232
n = 200	Mean	1.5560	0.5003	0.4811	0.3070	2.0658
	RB(%)	3.7369	0.0556	−3.7651	2.3506	3.2930
	MSE	0.0698	0.0010	0.0077	0.0093	0.0500
n = 400	Mean	1.5282	0.5001	0.4905	0.3034	2.0309
	RB(%)	1.8850	0.0398	−1.8974	1.1381	1.5470
	MSE	0.0350	0.0005	0.0038	0.0044	0.0226

Table 2. Fitted CMP-ARMA$(2,0)$, Negative Binomial GARMA$(2,0)$, and Poisson GARMA$(2,0)$ models for weekly number of hospitalizations data.

Model	CMLE	SE	p-value	AIC	BIC
CMP-ARMA	$\widehat{\alpha} = 2.6812$	0.3760	< 0.0001	2324.07	2341.68
	$\widehat{\beta}_1 = -0.1022$	0.0180	< 0.0001		
	$\widehat{\phi}_1 = 0.3504$	0.0631	< 0.0001		
	$\widehat{\phi}_2 = 0.1693$	0.0629	0.0071		
	$\widehat{\nu} = 0.3986$	0.0327	< 0.0001		
Negative Binomial	$\widehat{\beta}_0 = 5.5818$	0.0130	< 0.0001	2323.84	2341.45
	$\widehat{\beta}_1 = -0.1032$	0.0180	< 0.0001		
	$\widehat{\phi}_1 = 0.3516$	0.0633	< 0.0001		
	$\widehat{\phi}_2 = 0.1707$	0.0629	0.0067		
	$\widehat{\alpha} = 1.5049$	0.2249	< 0.0001		
Poisson GARMA	$\widehat{\beta}_0 = 5.5819$	0.0082	< 0.0001	2467.83	2481.91
	$\widehat{\beta}_1 = -0.1022$	0.0113	< 0.0001		
	$\widehat{\phi}_1 = 0.3504$	0.0398	< 0.0001		
	$\widehat{\phi}_2 = 0.1693$	0.0397	< 0.0001		

Table 3. Fitted CMP-ARMA$(1,1)$, Negative Binomial GARMA$(1,1)$, and Poisson GARMA$(1,1)$ models for pedestrians counts data.

Model	CMLE	SE	p-value	AIC	BIC
CMP-ARMA	$\widehat{\alpha} = 0.3217$	0.0457	< 0.0001	1246.89	1263.79
	$\widehat{\phi}_1 = 0.4511$	0.0537	< 0.0001		
	$\widehat{\theta}_1 = 0.2585$	0.0456	< 0.0001		
	$\widehat{\nu} = 2.4428$	0.1822	< 0.0001		
Negative Binomial	$\widehat{\beta}_0 = 0.5920$	0.0865	< 0.0001	1338.77	1355.67
	$\widehat{\phi}_1 = 0.4233$	0.0704	< 0.0001		
	$\widehat{\theta}_1 = 0.2671$	0.0644	< 0.0001		
	$\widehat{\sigma} = 0.0010$	0.1279	0.9937		
Poisson	$\widehat{\beta}_0 = 0.5911$	0.0794	< 0.0001	1336.86	1349.54
	$\widehat{\phi}_1 = 0.4257$	0.0700	< 0.0001		
	$\widehat{\theta}_2 = 0.2662$	0.0649	< 0.0001		