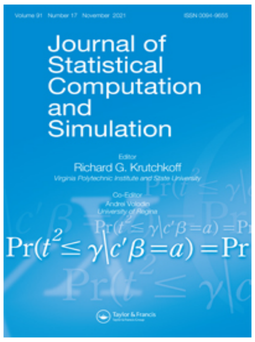

Journal of Statistical Computation and Simulation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Conway–Maxwell–Poisson seasonal autoregressive moving average model

Moizés da Silva Melo & Airlane Pereira Alencar

To cite this article: Moizés da Silva Melo & Airlane Pereira Alencar (2021): Conway–Maxwell–Poisson seasonal autoregressive moving average model, Journal of Statistical Computation and Simulation, DOI: 10.1080/00949655.2021.1955887

To link to this article: https://doi.org/10.1080/00949655.2021.1955887

Published online: 21 Jul 2021.

Submit your article to this journal

Article views: 65

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=gscs20

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
https://doi.org/10.1080/00949655.2021.1955887

Taylor & Francis
Taylor & Francis Group

Check for updates

Conway–Maxwell–Poisson seasonal autoregressive moving average model

Moizés da Silva Melo [a,b] and Airlane Pereira Alencar [a]

[a]Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, SP, Brazil; [b]Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil

ABSTRACT

This work proposes a new class of models, namely Conway–Maxwell–Poisson seasonal autoregressive moving average model (CMP-SARMA), which extends the class of Conway–Maxwell–Poisson autoregressive moving average models by including seasonal components to the dynamic model structure. The proposed class of models assumes a Conway–Maxwell–Poisson conditional distribution for the response variable, which allows us to model univariate time series of non-negative counts with overdispersion, equidispersion, and underdispersion. We estimated the parameters by conditional maximum likelihood. We also present closed-form expressions for the conditional score function and conditional Fisher information matrix. In addition, hypothesis testing, diagnostic analysis, and forecasting are proposed and asymptotic results are discussed. A Monte Carlo simulation study is conducted to evaluate the finite sample properties of the estimators. Finally, we present an application of the new model to real data and compare the results with other models in the literature.

ARTICLE HISTORY
Received 16 March 2021
Accepted 11 July 2021

KEYWORDS
CMP-SARMA; Conway–Maxwell–Poisson distribution; time series of counts; seasonality

1. Introduction

Recently, the generalized autoregressive moving average (GARMA) model proposed by [1] has been considered in some time series applications/studies [2–5]. The GARMA model extends the univariate Gaussian autoregressive moving average (ARMA) model to a flexible observation driven model for non-Gaussian time series data. Similar to the generalized linear model (GLM) introduced by [6], the conditional mean of the response variable is modelled directly by a regression structure through a link function.

The model proposed by [1] assumes that the conditional distribution of the dependent variable belongs to the exponential family given the process history. In the same approach, [7] developed dynamic models in the beta distribution family (βARMA), [8] introduced a dynamic class of models taking values in the double bounded interval following the Kumaraswamy distribution, and [9] proposed a dynamic class of models for random variables belonging to the class of symmetric distributions. Recently, [10] developed a dynamic regression model based on the Conway–Maxwell–Poisson (CMP) distribution for the

CONTACT Moizés da Silva Melo moizes.melo@ufrn.br, moizessm@ime.usp.br

© 2021 Informa UK Limited, trading as Taylor & Francis Group

analysis of time series of counts with equidispersion, underdispersion, and overdispersion. Although the mentioned models can be used for time series with seasonality, using sine/cosine functions as covariates, such a strategy is not appropriate when the seasonality is stochastic [11].

Modelling seasonal time series has been the focus of extensive research in the literature. In practical situations, the well-known seasonal autoregressive integrated moving average (SARIMA) model [12] has been frequently used for modelling univariate time series. However, many real data often do not adhere to the assumption of normality required by this model [13]. Some works have shown an increasing interest in non-Gaussian seasonal time series models, such as [11,14,15], and [16]. However, the study of seasonal time series of counts with equidispersion, underdispersion, and overdispersion has received less attention in the literature. Thus, this paper aims to give a contribution towards this direction.

Based on the above discussion, we propose a class of Conway–Maxwell–Poisson seasonal autoregressive moving average (CMP-SARMA) models, which extends the Conway–Maxwell–Poisson autoregressive moving average (CMP-ARMA) and is capable to model seasonal time series of count data that exhibit overdispersion, equidispersion, and underdispersion. For this purpose, we adopted the mean-parametrized CMP distribution proposed by [17] that allows the mean to be modelled directly. The model parameters are estimated using the conditional maximum likelihood estimation. In addition, some residual and diagnostic tools are proposed and discussed. The performance in small samples of the estimators is evaluated via Monte Carlo simulations

This paper is organized as follows. Initially, we present the proposed model in Section 2. The parameters of the model are estimated by the conditional maximum likelihood method in Section 3, where we also provide closed-form expressions for the conditional score vector and conditional Fisher information matrix. The discussion in Section 3 also provides confidence intervals and hypothesis testing. Section 4 gives some diagnostic measures and forecasting. In Section 5, we conduct a brief simulation study. Section 6 presents an empirical application to illustrate the proposed model. Finally, Section 7 presents some conclusions.

2. The proposed model

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a sequence of random variables and let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ be the corresponding r-dimensional vector of explanatory variables. In the CMP-SARMA model, conditional on its past, each observation Y_t follows the CMP distribution with mean μ_t and dispersion ν, that is, the conditional probability function is given by

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

where $\mathcal{F}_{t-1} = \{Y_{t-1},\ldots,Y_1,\boldsymbol{X}_{t-1},\ldots,\boldsymbol{X}_1\}$ is the σ-field generated by the past values of the series and of the covariate data, $\lambda(\mu_t,\nu)$ is a function of μ_t and ν, given by the solution to

$$0 = \sum_{s=0}^{\infty}(s-\mu_t)\frac{\lambda^s}{(s!)^{\nu}} \tag{2}$$

and

$$Z(\lambda(\mu_t,\nu),\nu)=\sum_{s=0}^{\infty}\frac{\lambda(\mu_t,\nu)^s}{(s!)^{\nu}} \tag{3}$$

is the normalizing constant. Furthermore, ν is the dispersion parameter such that $\nu > 1$ implies underdispersion and $\nu < 1$ implies overdispersion relative to a Poisson distribution with same mean. The normalizing constant $Z(\lambda(\mu_t,\nu),\nu)$ and the parameter $\lambda(\mu_t,\nu)$ presented above can be numerically obtained by truncating the infinite sums given in Equations (2) and (3), respectively. The conditional mean of Y_t is given by $E(Y_t \mid \mathcal{F}_{t-1}) = \mu_t$.

The proposed Conway–Maxwell–Poisson seasonal autoregressive moving average model, CMP-SARMA$(p,q)\times(P,Q)_S$, can be written as

$$\phi(B)\Phi(B^S)[g(y_t)-x_t^{\top}\boldsymbol{\beta}]=\theta(B)\Theta(B^S)r_t, \tag{4}$$

where $\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_l)^{\top}$ is the $(l+1)$-dimensional vector of unknown parameters, $\boldsymbol{x}_t=(1,x_{1,t},\ldots,x_{l,t})^{\top}$ is the $(l+1)$-dimensional vector containing the covariates at time t, $r_t=g(y_t)-g(\mu_t)$ is the uncorrelated random errors sequence measurable with respect to $\mathcal{F}_t$, $g(\cdot)$ is a link function, $\phi(B)$ is the non-seasonal autoregressive polynomial, $\theta(B)$ is the non-seasonal moving average polynomial, $\Phi(B^S)$ is the seasonal autoregressive polynomial, and $\Theta(B^S)$ is the seasonal moving average polynomial, defined by

$$\begin{aligned}
\phi(B)&=1-\phi_1B-\phi_2B^2-\cdots-\phi_pB^p,\\
\theta(B)&=1-\theta_1B-\theta_2B^2-\cdots-\theta_qB^q,\\
\Phi(B^S)&=1-\Phi_1B^S-\Phi_2B^{2S}-\cdots-\Phi_PB^{SP},\\
\Theta(B^S)&=1-\Theta_1B^S-\Theta_2B^{2S}-\cdots-\Theta_QB^{SQ},
\end{aligned}$$

where p, q, P, and Q are the orders of the respective polynomials, S is the seasonal period, and B is a backshift operator, that is $B^d y_t = y_{t-d}$.

We can rewrite Equation (4) as

$$\begin{aligned}
\eta_t=g(\mu_t)&=x_t^{\top}\boldsymbol{\beta}+\sum_{i=1}^{p}\phi_i[g(y_{t-i})-x_{t-i}^{\top}\boldsymbol{\beta}]+\sum_{I=1}^{P}\Phi_I[g(y_{t-IS})-x_{t-IS}^{\top}\boldsymbol{\beta}]\\
&-\sum_{i=1}^{p}\sum_{I=1}^{P}\phi_i\Phi_I[g(y_{t-(i+IS)})-x_{t-(i+IS)}^{\top}\boldsymbol{\beta}]-\sum_{j=1}^{q}\theta_j r_{t-j}-\sum_{J=1}^{Q}\Theta_J r_{t-(JS)}\\
&+\sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J r_{t-(j+JS)},
\end{aligned} \tag{5}$$

where $\eta_t=g(\mu_t)$ is the linear predictor. The structure of the model is the same as in [16]. Similarly to the SARMA model, the transformed mean $(g(\mu_t))$ may depend on lagged observations $(g(y_{t-i}))$ and errors r_{t-i}, for lags $i=1,2,\ldots$; and/or seasonal lags $i=1S,2S,\ldots$. Choosing an identity link function $(g(\mu_t)=\mu_t)$ implies some restrictions to ensure $\mu_t\geq 0$. Thus, to circumvent this problem, we chose the logarithmic link function.

Also, to avoid calculating the logarithm of observations equal to zero, we replace $y_{t-j} = 0$ in Equation (4) with $y^*_{t-j} = max(y_{t-j}, c)$, where c is an arbitrary small value, such that $0 < c < 1$.

The CMP-SARMA model extends the approach proposed by [10] by incorporating a seasonal autoregressive moving average (SARMA) structure.

3. Parameter estimation

The parameter estimation of the CMP-SARMA model can be obtained by the maximum conditional likelihood method based on the first m observations, where $m = \max(p + PS, q + QS)$.

Let $y_1, \ldots, y_t, t = 1, \ldots, n$, be a time series with length n. To fit a CMP-SARMA$(p, q) \times (P, Q)_S$ model, let $\boldsymbol{\gamma} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\Phi}^\top, \boldsymbol{\Theta}^\top, \nu)^\top$ be the regression parameter vector. The conditional log-likelihood function is given by

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

where

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

3.1. Conditional score vector

The conditional score vector is given by taking first derivatives of the conditional log-likelihood function given in (6) with respect to each element of $\boldsymbol{\gamma}$. By the chain rule, for $\gamma_i \neq \nu$, $i = 1, 2, \ldots, l + p + q + P + Q + 1$, we have

$$U_{\gamma_i}(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_i} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \nu)}{\partial \lambda(\mu_t, \nu)} \frac{\partial \lambda(\mu_t, \nu)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_i}. \tag{7}$$

Observe that

$$\frac{\partial \ell_t(\mu_t, \nu)}{\partial \lambda(\mu_t, \nu)} = \frac{y_t - \mu_t}{\lambda(\mu_t, \nu)}, \quad \frac{\partial \lambda(\mu_t, \nu)}{\partial \mu_t} = \frac{\lambda(\mu_t, \nu)}{V(\mu_t, \nu)}, \quad \text{and} \quad \frac{d\mu_t}{d\eta_t} = \frac{1}{g'(\mu_t)},$$

where $V(\mu_t, \nu) = \sum_{y_t=0}^{\infty} \frac{(y_t - \mu_t)^2 \lambda_t(\mu_t, \nu)^y_t}{(y_t!)^\nu Z(\lambda_t(\mu_t, \nu), \nu)}$ is the variance of Y_t. Substituting these results in (7), we obtain

$$U_{\gamma_i}(\boldsymbol{\gamma}) = \sum_{t=1}^{n} \frac{y_t - \mu_t}{V(\mu_t, \nu) g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_i}, \quad \text{for } \gamma_i \notin \nu. \tag{8}$$

We also have

$$\frac{\partial \eta_t}{\partial \phi_i} = [g(y_{t-i}) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta}] \Phi(B^S) + \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \phi_i} + \sum_{J=1}^{Q} \Theta_J \frac{\partial \eta_{t-JS}}{\partial \phi_i}$$

$$- \sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J\frac{\partial\eta_{t-(j+JS)}}{\partial\phi_i},$$

$$\frac{\partial\eta_t}{\partial\theta_j} = -r_{t-j}\Theta(B^S) + \sum_{i=1}^{q}\theta_i\frac{\partial\eta_{t-i}}{\partial\theta_j} + \sum_{J=1}^{Q}\Theta_J\frac{\partial\eta_{t-JS}}{\partial\theta_j} - \sum_{i=1}^{q}\sum_{J=1}^{Q}\theta_i\Theta_J\frac{\partial\eta_{t-(i+JS)}}{\partial\theta_j},$$

$$\frac{\partial\eta_t}{\partial\Phi_I} = [g(y_{t-IS}) - \boldsymbol{x}_{t-IS}^\top\boldsymbol{\beta}]\phi(B) + \sum_{j=1}^{q}\theta_j\frac{\partial\eta_{t-j}}{\partial\Phi_I} + \sum_{J=1}^{Q}\Theta_J\frac{\partial\eta_{t-JS}}{\partial\Phi_I}$$

$$- \sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J\frac{\partial\eta_{t-(j+JS)}}{\partial\Phi_I},$$

$$\frac{\partial\eta_t}{\partial\Theta_J} = -r_{t-JS}\theta(B) + \sum_{j=1}^{q}\theta_j\frac{\partial\eta_{t-j}}{\partial\Theta_J} + \sum_{I=1}^{Q}\Theta_I\frac{\partial\eta_{t-IS}}{\partial\Theta_J} - \sum_{j=1}^{q}\sum_{I=1}^{Q}\theta_j\Theta_I\frac{\partial\eta_{t-(j+IS)}}{\partial\Theta_J},$$

$$\frac{\partial\eta_t}{\partial\beta_k} = x_{tk} - \sum_{i=1}^{p}\phi_i x_{(t-i)k} - \sum_{I=1}^{P}\Phi_I x_{(t-IS)k} + \sum_{i=1}^{p}\sum_{I=1}^{P}\phi_i\Phi_I x_{(t-(i+IS))k} + \sum_{j=1}^{q}\theta_j\frac{\partial\eta_{t-j}}{\partial\beta_k}$$

$$+ \sum_{J=1}^{Q}\Theta_J\frac{\partial\eta_{t-JS}}{\partial\beta_k} - \sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J\frac{\partial\eta_{t-(j+JS)}}{\partial\beta_k}.$$

Let $\boldsymbol{y} = (y_{m+1},\ldots,y_n)^\top$, $\boldsymbol{\mu} = (\mu_{m+1},\ldots,\mu_n)^\top$, $\boldsymbol{T} = \text{diag}\{1/g'(\mu_{m+1}),\ldots,1/g'(\mu_n)\}$, and $\boldsymbol{V} = \text{diag}\{1/V(\mu_{m+1},\nu),\ldots,1/V(\mu_n,\nu)\}$. Also, let $\boldsymbol{Z}$, $\boldsymbol{A}$, $\mathcal{A}$, $\boldsymbol{M}$, and $\mathcal{M}$ be the matrices with dimension $(n-m)\times(r+1)$, $(n-m)\times p$, $(n-m)\times P$, $(n-m)\times q$, and $(n-m)\times Q$, respectively, for which the (i,j)-th elements are given by

$$Z_{i,j} = \frac{\partial\eta_{m+i}}{\partial\beta_j}, \quad A_{i,j} = \frac{\partial\eta_{m+i}}{\partial\phi_j}, \quad \mathcal{A}_{i,j} = \frac{\partial\eta_{m+i}}{\partial\Phi_j},$$
$$M_{i,j} = \frac{\partial\eta_{m+i}}{\partial\theta_j}, \quad \text{and} \quad \mathcal{M}_{i,j} = \frac{\partial\eta_{m+i}}{\partial\Theta_j}.$$

For each $\gamma_i \notin \nu$ in Equation (8), each element can be written in matrix form as

$$\boldsymbol{U}_\beta(\boldsymbol{\gamma}) = \boldsymbol{Z}^\top\boldsymbol{T}\boldsymbol{V}(\boldsymbol{y}-\boldsymbol{\mu}), \quad \boldsymbol{U}_\phi(\boldsymbol{\gamma}) = \boldsymbol{A}^\top\boldsymbol{T}\boldsymbol{V}(\boldsymbol{y}-\boldsymbol{\mu}),$$
$$\boldsymbol{U}_\Phi(\boldsymbol{\gamma}) = \mathcal{A}^\top\boldsymbol{T}\boldsymbol{V}(\boldsymbol{y}-\boldsymbol{\mu}), \quad \boldsymbol{U}_\theta(\boldsymbol{\gamma}) = \boldsymbol{M}^\top\boldsymbol{T}\boldsymbol{V}(\boldsymbol{y}-\boldsymbol{\mu}),$$
$$\boldsymbol{U}_\Theta(\boldsymbol{\gamma}) = \mathcal{M}^\top\boldsymbol{T}\boldsymbol{V}(\boldsymbol{y}-\boldsymbol{\mu}).$$

The derivative of $\ell(\boldsymbol{\gamma})$ with respect to the dispersion parameter ν is given by

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

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

The conditional score vector can then be written as

$$U(\boldsymbol{\gamma}) = (\boldsymbol{U}_\beta(\boldsymbol{\gamma})^\top, \boldsymbol{U}_\phi(\boldsymbol{\gamma})^\top, \boldsymbol{U}_\theta(\boldsymbol{\gamma})^\top, \boldsymbol{U}_\Phi(\boldsymbol{\gamma})^\top, \boldsymbol{U}_\Theta(\boldsymbol{\gamma})^\top, U_\nu(\boldsymbol{\gamma})).$$

The conditional maximum likelihood estimators (CMLE) are obtained from the solution of the system of nonlinear equations $\boldsymbol{U}(\boldsymbol{\gamma}) = \mathbf{0}$, where $\mathbf{0}$ is a vector of zeros with dimension $(l+p+q+P+Q+2)$. Such a system does not have an analytical solution, being necessary to apply iterative numerical methods. Here, we apply the `nlminb` optimization algorithm [18] available from the stats package in `R` software [19]. We assume the initial values for η_t and r_t to be equal zero, both for $t = 1, 2, \ldots, m$. Next, we shall obtain η_t and r_t for $t > m$ recursively using (5).

3.2. Conditional information matrix

This section provides analytic formulae for the conditional Fisher information matrix, which will be used later to construct the asymptotic confidence intervals and hypothesis tests. The conditional Fisher information matrix for $\boldsymbol{\gamma}$ is given by $K(\boldsymbol{\gamma}) = \mathbb{E}[-\partial^2\ell(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top]$, which requires the expectations of the second derivatives of the conditional log-likelihood function.

The second order derivatives of the log-likelihood function, for $i, j \in \{1, \ldots, l+p+q+P+Q+1\}$ $(\gamma_i \neq \nu,)$, are given by

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

Since $E(\frac{\partial\ell_t(\mu_t,\nu)}{\partial\mu_t}|\mathcal{F}_{t-1}) = 0$ the expected value of the derivative above is given by

$$E\left(\frac{\partial^2\ell(\boldsymbol{\gamma})}{\partial\gamma_i\partial\gamma_j}\middle|\mathcal{F}_{t-1}\right) = \sum_{t=m+1}^{n} E\left(\frac{\partial^2\ell_t(\mu_t,\nu)}{\partial\mu_t^2}\middle|\mathcal{F}_{t-1}\right)\left(\frac{\mathrm{d}\mu_t}{\mathrm{d}\eta_t}\right)^2\frac{\partial\eta_t}{\partial\gamma_i}\frac{\partial\eta_t}{\partial\gamma_j}. \tag{9}$$

The second order derivatives of $\ell_t(\mu_t, \nu)$ with respect to μ_t are given by

$$\frac{\partial^2\ell_t(\mu_t,\nu)}{\partial\mu_t^2} = \frac{-V(\mu_t,\nu) - (y_t-\mu_t)\,[m_3(\mu_t,\nu)/V(\mu_t,\nu)]}{V(\mu_t,\nu)^2},$$

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

$$E\left(\frac{\partial^2\ell_t(\mu_t,\nu)}{\partial\mu_t^2}\middle|\mathcal{F}_{t-1}\right) = \frac{-1}{V(\mu_t,\nu)}. \tag{10}$$

By replacing (10) in (9), it follows that

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

The second order derivatives of the log-likelihood function related to ν are given by [10]

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \nu \partial \gamma_j} &= \frac{\partial}{\partial \gamma_j}\left\{\sum_{t=m+1}^{n} A(\mu_t,\nu)\frac{(y_t-\mu_t)}{V(\mu_t,\nu)} - \left[\log(y_t!) - B(\mu_t,\nu)\right]\right\} \\ &= \sum_{t=m+1}^{n}\left\{\left[\frac{B(\mu_t,\nu)}{V(\mu_t,\nu)} + \frac{D(\mu_t,\nu)}{V(\mu_t,\nu)^2} - \frac{m_3(\mu_t,\nu)}{V(\mu_t,\nu)^3}\right]\frac{\partial \mu_t}{\partial \gamma_j}(y_t-\mu_t)\right\}\end{aligned}$$

and

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

where $C(\mu_t,\nu) = V_{\mu_t,\nu}(\log(y_t))$ and $D(\mu_t,\nu) = E_{\mu_t,\nu}[\log(y_t!)(y_t-\mu_t)^2]$.

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

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

and

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

Notice that ν is orthogonal to the other parameters.

Let $W = \text{diag}\{w_1,\ldots,w_n\}$, where

$$w_t = -\frac{A(\mu_t,\nu)^2}{V(\mu_t,\nu)} + C(\mu_t,\nu).$$

The conditional Fisher information matrix for $\boldsymbol{\gamma}$ is given by

$$\boldsymbol{K}(\boldsymbol{\gamma}) = \begin{pmatrix} K_{(\boldsymbol{\beta},\boldsymbol{\beta})} & \boldsymbol{K}_{(\boldsymbol{\beta},\boldsymbol{\phi})} & \boldsymbol{K}_{(\boldsymbol{\beta},\boldsymbol{\theta})} & \boldsymbol{K}_{(\boldsymbol{\beta},\boldsymbol{\Phi})} & \boldsymbol{K}_{(\boldsymbol{\beta},\boldsymbol{\Theta})} & K_{(\boldsymbol{\beta},\nu)} \\ \boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\beta})} & \boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\phi})} & \boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\theta})} & \boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\Phi})} & \boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\Theta})} & \boldsymbol{K}_{(\boldsymbol{\phi},\nu)} \\ \boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\beta})} & \boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\phi})} & \boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\theta})} & \boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\Phi})} & \boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\Theta})} & \boldsymbol{K}_{(\boldsymbol{\theta},\nu)} \\ \boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\beta})} & \boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\phi})} & \boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\theta})} & \boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\Phi})} & \boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\Theta})} & \boldsymbol{K}_{(\boldsymbol{\Phi},\nu)} \\ \boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\beta})} & \boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\phi})} & \boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\theta})} & \boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\Phi})} & \boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\Theta})} & \boldsymbol{K}_{(\boldsymbol{\Theta},\nu)} \\ K_{(\nu,\boldsymbol{\beta})} & \boldsymbol{K}_{(\nu,\boldsymbol{\phi})} & \boldsymbol{K}_{(\nu,\boldsymbol{\theta})} & \boldsymbol{K}_{(\nu,\boldsymbol{\Phi})} & \boldsymbol{K}_{(\nu,\boldsymbol{\Theta})} & K_{(\nu,\nu)} \end{pmatrix}, \tag{11}$$

where

$$
\begin{aligned}
&\boldsymbol{K}_{(\boldsymbol{\beta},\boldsymbol{\beta})} = \boldsymbol{Z}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{Z}, && \boldsymbol{K}_{(\boldsymbol{\beta},\boldsymbol{\phi})} = \boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\beta})}^\top = \boldsymbol{Z}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{A},\\
&\boldsymbol{K}_{(\boldsymbol{\beta},\boldsymbol{\theta})} = \boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\beta})}^\top = \boldsymbol{Z}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{M}, && \boldsymbol{K}_{(\boldsymbol{\beta},\boldsymbol{\Phi})} = \boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\beta})}^\top = \boldsymbol{Z}^\top \boldsymbol{V}\boldsymbol{T}^2\mathcal{A},\\
&\boldsymbol{K}_{(\boldsymbol{\beta},\boldsymbol{\Theta})} = \boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\beta})}^\top = \boldsymbol{Z}^\top \boldsymbol{V}\boldsymbol{T}^2\mathcal{M}, && \boldsymbol{K}_{(\boldsymbol{\beta},\nu)} = \boldsymbol{K}_{(\nu,\boldsymbol{\beta})}^\top = \boldsymbol{0},\\
&\boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\phi})} = \boldsymbol{A}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{A}, && \boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\theta})} = \boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\phi})}^\top = \boldsymbol{A}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{M},\\
&\boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\Phi})} = \boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\phi})}^\top = \boldsymbol{A}^\top \boldsymbol{V}\boldsymbol{T}^2\mathcal{A}, && \boldsymbol{K}_{(\boldsymbol{\phi},\boldsymbol{\Theta})} = \boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\phi})}^\top = \boldsymbol{A}^\top \boldsymbol{V}\boldsymbol{T}^2\mathcal{M},\\
&\boldsymbol{K}_{(\boldsymbol{\phi},\nu)} = \boldsymbol{K}_{(\nu,\boldsymbol{\phi})}^\top = \boldsymbol{0}, && \boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\theta})} = \boldsymbol{M}^\top \boldsymbol{V}\boldsymbol{T}^2\boldsymbol{M},\\
&\boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\Phi})} = \boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\theta})}^\top = \boldsymbol{M}^\top \boldsymbol{V}\boldsymbol{T}^2\mathcal{A}, && \boldsymbol{K}_{(\boldsymbol{\theta},\boldsymbol{\Theta})} = \boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\theta})}^\top = \boldsymbol{M}^\top \boldsymbol{V}\boldsymbol{T}^2\mathcal{M},\\
&\boldsymbol{K}_{(\boldsymbol{\theta},\nu)} = \boldsymbol{K}_{(\nu,\boldsymbol{\theta})}^\top = \boldsymbol{0}, && \boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\Phi})} = \mathcal{A}^\top \boldsymbol{V}\boldsymbol{T}^2\mathcal{A},\\
&\boldsymbol{K}_{(\boldsymbol{\Phi},\boldsymbol{\Theta})} = \boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\Phi})}^\top = \mathcal{A}^\top \boldsymbol{V}\boldsymbol{T}^2\mathcal{M}, && \boldsymbol{K}_{(\boldsymbol{\Phi},\nu)} = \boldsymbol{K}_{(\nu,\boldsymbol{\Phi})}^\top = \boldsymbol{0},\\
&\boldsymbol{K}_{(\boldsymbol{\Theta},\boldsymbol{\Theta})} = \mathcal{M}^\top \boldsymbol{V}\boldsymbol{T}^2\mathcal{M}, && K_{(\boldsymbol{\Theta},\nu)} = K_{(\nu,\boldsymbol{\Theta})}^\top = \boldsymbol{0},\\
&K_{(\nu,\nu)} = \mathrm{tr}(\boldsymbol{W}).
\end{aligned}
$$

As normalizing the constant in (3), the series $A(\mu_t,\nu)$, $B(\mu_t,\nu)$, $C(\mu_t,\nu)$ and $D(\mu_t,\nu)$ can be computed by truncating the numerical series. Under the usual regularity conditions and for n sufficiently large, the conditional maximum likelihood estimator $\widehat{\boldsymbol{\gamma}}$ of the parameter vector $\boldsymbol{\gamma}$ is asymptotically normally distributed [8,20,21], that is,

$$
\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\phi}} \\ \widehat{\boldsymbol{\Phi}} \\ \widehat{\boldsymbol{\theta}} \\ \widehat{\boldsymbol{\Theta}} \\ \widehat{\nu} \end{pmatrix} \sim N_{r+p+q+P+Q+2}\left(\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\phi} \\ \boldsymbol{\Phi} \\ \boldsymbol{\theta} \\ \boldsymbol{\Theta} \\ \nu \end{pmatrix}, \boldsymbol{K}(\boldsymbol{\gamma})^{-1} \right), \tag{12}
$$

where $N_{l+p+q+P+Q+2}$ denotes the $(l+p+q+P+Q+2)$-dimensional normal distribution, $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}$, and $\widehat{\nu}$ are the CMLE of $\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\Phi}, \boldsymbol{\theta}, \boldsymbol{\Theta}$, and ν, respectively, and $\boldsymbol{K}(\widehat{\boldsymbol{\gamma}})^{-1}$ is the conditional Fisher information inverse matrix.

3.3. Confidence intervals and hypothesis testing

Consider the null hypothesis $\mathcal{H}_0 : \gamma_i = \gamma_i^0$ versus $\mathcal{H}_1 : \gamma_i \neq \gamma_i^0$, where γ_i^0 is a specified value for the unknown parameter γ_i. A useful statistic that is particularly convenient to test individual parameters [22] is the so-called z statistic, which is given by

$$
z = \frac{\widehat{\gamma_i} - \gamma_i}{\sqrt{k^{ii}}}, \tag{13}
$$

where k^{ii} is the i-th diagonal element of $\boldsymbol{K}(\widehat{\boldsymbol{\gamma}})^{-1}$.

Under $\mathcal{H}_0$ and for large n, z follows approximately a standard normal distribution. More general hypothesis testing inference can also be performed using the log-partial likelihood ratio, Wald, and score statistics. Under $\mathcal{H}_0$, all the mentioned test statistics converge to a χ^2 distribution. See [23] for further details.

We can also obtain asymptotic confidence intervals for each parameter γ_i. An approximate $100(1-\alpha)\%$ confidence interval for γ_i 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 distribution function of the standardized normal distribution $N(0,1)$.

4. Diagnostic measures and forecasting

This section introduces some model selection criteria and procedures to test the adequacy and goodness-of-fit of the proposed model. For the model selection, we adopted the following two information criteria: the Akaike Information Criterion (AIC) [24] and the Schwarz Information Criterion (SIC) [25], given, respectively, by

$$\begin{aligned} \text{AIC} &= -\widehat{\ell} + 2k, \\ \text{SIC} &= -\widehat{\ell} + \log(n)k, \end{aligned} \tag{14}$$

where $k = l+p+q+P+Q+2$ is the number of parameters in the model.

Residual analysis is an important technique for checking model adequacy. Here, we consider the randomized quantile residuals introduced by [26]. When the response variable is discrete, the randomized quantile residuals are given by $r_t^{(q)} = \Phi^{-1}(u_t)$, where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution, u_i is a random variable uniformly distributed in the interval $[F(y_t - 1 \mid \mathcal{F}_{t-1}), F(y_t \mid \mathcal{F}_{t-1})]$, and $F(\cdot)$ is the cumulative distribution function of the observations.

To test the validity of the assumed distribution in the proposed model, we also use a non-randomized yet uniform version of the probability integral transformation (PIT) proposed by [27] for time series models for count data. It utilizes the conditional cumulative distribution function given the observed count y_t via

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

The calibration can be assessed by comparing the mean PIT, defined by

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

to the cumulative distribution function of a uniform random variable. This comparison can be performed empirically by plotting a non-randomized PIT histogram at J equidistant bins, where the height of the histogram is $f_j = \overline{F}(j/J) - \overline{F}((j-1)/J)$ for $j = 1, \ldots, J$, and check for uniformity [28,29]. U-shaped and inverse U-shaped histograms indicate, respectively, underdispersed and overdispersion predictive distributions.

We can obtain h-steps ahead forecasts for the CMP-SARMA model as follows

$$\widehat{y}_n(h) = \exp\left(\widehat{\alpha} + \sum_{i=1}^{p} \widehat{\phi}_i[g(y_{n+h-i})] + \sum_{I=1}^{P} \widehat{\Phi}_I[g(y_{n+h-IS})] - \sum_{i=1}^{p}\sum_{I=1}^{P} \widehat{\phi}_i\widehat{\Phi}_I[g(y_{n+h-(i+IS)})] - \sum_{j=1}^{q} \widehat{\theta}_j[\widehat{r}_{n+h-j}] - \sum_{J=1}^{Q} \widehat{\Theta}_J[\widehat{r}_{n+h-JS}] + \sum_{j=1}^{q}\sum_{J=1}^{Q} \widehat{\theta}_j\widehat{\Theta}_J[\widehat{r}_{n+h-(j+JS)}]\right),$$

where

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

5. Monte Carlo simulation study

In what follows, we shall evaluate the asymptotic properties of the CMLE for the proposed model through a Monte Carlo (MC) simulation study using the CMP-SARMA$(1,1) \times (1,1)_{12}$ model with three different values for the dispersion parameter: $\nu \in \{0.5, 1.0, 2.0\}$. The systematic component of the model is given by

$$\log(\mu_t) = \phi_1[\log(y_{t-1}) - \beta_0] + \Phi_1[\log(y_{t-12}) - \beta_0] - \phi_1\Phi_1[\log(y_{t-13}) - \beta_0] - \theta_1[r_{t-1}] - \Theta_1[r_{t-12}] + \theta_1\Theta_1 r_{t-13}, \quad t = 14, \ldots, n,$$

where $\beta_0 = 2.0$, $\phi_1 = 0.5$, $\theta_1 = -0.4$, $\Phi_1 = -0.2$, and $\Theta_1 = 0.3$. All computational routines developed in this paper were implemented and performed in `R software` [19]. We generate 5000 Monte Carlo replicates of each experiment with $n \in \{100, 100, 200, 400, 800\}$. For each experiment, we evaluate the mean, percentage relative bias (RB %), defined as $\{E(\widehat{\theta}) - \theta\}/\theta$, and mean squared error (MSE). After a preliminary study on the parameter estimates, we truncate all infinite sums in this paper at the 1001st term.

The results for all scenarios are shown in Table 1. We note that the bias decreases and that the MSE tends toward zero as the size of the sample increases, indicating the consistency property of the CMLE. These results indicate that the CMLE appeared to perform well. We also note that the seasonal estimators present a larger relative bias in all scenarios, as expected, since we have less information about the seasonal parameters in the observations, because we need 12 observations to complete a seasonal cycle. Such fact was also verified for the βSARMA model in [16].

6. Empirical application

In this section, for illustrative purposes, we analyse and discuss a real application. We fitted the CMP-SARMA model to monitor the number of robberies in the City of Chicago from January 2020 to April 2020. This dataset was obtained from the Chicago Police Department[1] and contains the hourly number of robberies committed in the City of

Table 1. Monte Carlo simulation results of the CMLE for the CMP-ARMA $(1,1) \times (1,1)_{12}$ model.

		Scenario 1 – overdispersion					
	Parameters	α 2.0	ϕ_1 0.5	Φ_1 −0.2	θ_1 −0.4	Θ_1 0.3	ν 0.5
	Mean	2.299	0.469	−0.300	−0.413	0.211	0.539
n = 100	RB(%)	14.966	−6.113	50.058	3.366	−29.501	7.937
	MSE	0.506	0.018	0.040	0.019	0.055	0.018
	Mean	2.161	0.486	−0.261	−0.406	0.249	0.518
n = 200	RB(%)	8.043	−2.830	30.386	1.538	−16.880	3.566
	MSE	0.203	0.007	0.021	0.008	0.024	0.007
	Mean	2.115	0.490	−0.244	−0.404	0.266	0.512
n = 300	RB(%)	5.746	−2.045	22.012	0.938	−11.344	2.367
	MSE	0.122	0.005	0.014	0.005	0.015	0.005
	Mean	2.089	0.493	−0.237	−0.402	0.273	0.510
n = 400	RB(%)	4.452	−1.387	18.260	0.595	−9.075	2.015
	MSE	0.090	0.003	0.011	0.004	0.012	0.004
	Mean	2.055	0.497	−0.225	−0.398	0.289	0.509
n = 800	RB(%)	2.750	−0.685	12.707	−0.392	−3.732	1.904
	MSE	0.044	0.002	0.007	0.002	0.006	0.003
		Scenario 2 – equidispersion					
	Parameters	α 2.0	ϕ_1 0.5	Φ_1 −0.2	θ_1 −0.4	Θ_1 0.3	ν 2.0
	Mean	2.298	0.469	−0.299	−0.414	0.214	1.079
n = 100	RB(%)	14.924	−6.211	49.434	3.500	−28.668	7.883
	MSE	0.513	0.018	0.041	0.020	0.058	0.037
	Mean	2.159	0.485	−0.259	−0.407	0.251	1.034
n = 200	RB(%)	7.958	−2.918	29.292	1.775	−16.431	3.410
	MSE	0.207	0.007	0.021	0.008	0.025	0.013
	Mean	2.112	0.489	−0.241	−0.405	0.266	1.020
n = 300	RB(%)	5.599	−2.163	20.261	1.334	−11.369	2.022
	MSE	0.125	0.005	0.013	0.005	0.015	0.008
	Mean	2.085	0.492	−0.232	−0.404	0.273	1.015
n = 400	RB(%)	4.272	−1.539	16.036	1.038	−9.166	1.545
	MSE	0.092	0.003	0.010	0.004	0.011	0.006
	Mean	2.047	0.496	−0.218	−0.402	0.285	1.007
n = 800	RB(%)	2.347	−0.802	9.239	0.464	−5.079	0.704
	MSE	0.042	0.002	0.005	0.002	0.005	0.003
		Scenario 3 – underdispersion					
	Parameters	α 2.0	ϕ_1 0.5	Φ_1 −0.2	θ_1 −0.4	Θ_1 0.3	ν 2.0
	Mean	2.300	0.468	−0.298	−0.415	0.218	2.158
n = 100	RB(%)	14.998	−6.360	48.870	3.677	−27.260	7.886
	MSE	0.525	0.019	0.042	0.021	0.060	0.146
	Mean	2.159	0.485	−0.257	−0.407	0.254	2.068
n = 200	RB(%)	7.944	−3.008	28.688	1.866	−15.394	3.422
	MSE	0.212	0.008	0.021	0.008	0.025	0.053
	Mean	2.110	0.489	−0.239	−0.406	0.269	2.041
n = 300	RB(%)	5.518	−2.232	19.354	1.386	−10.371	2.061
	MSE	0.129	0.005	0.014	0.005	0.015	0.033
	Mean	2.083	0.492	−0.230	−0.404	0.275	2.032
n = 400	RB(%)	4.154	−1.560	15.243	1.067	−8.354	1.575
	MSE	0.094	0.004	0.010	0.004	0.012	0.023
	Mean	2.046	0.496	−0.217	−0.402	0.286	2.015
n = 800	RB(%)	2.276	−0.815	8.724	0.493	−4.571	0.730
	MSE	0.044	0.002	0.005	0.002	0.005	0.011

Table 2. Chicago robbery data summary statistics aggregated by 4-hour periods.

Period (hour interval)	Mean	Variance	Minimum	Maximum
00:00 03:59	2.79	3.12	0	8
04:00 07:59	1.61	2.21	0	8
08:00 11:59	2.01	2.21	0	8
12:00 15:59	3.71	4.11	0	10
16:00 19:59	4.72	6.30	0	13
20:00 23:59	4.28	5.79	0	13

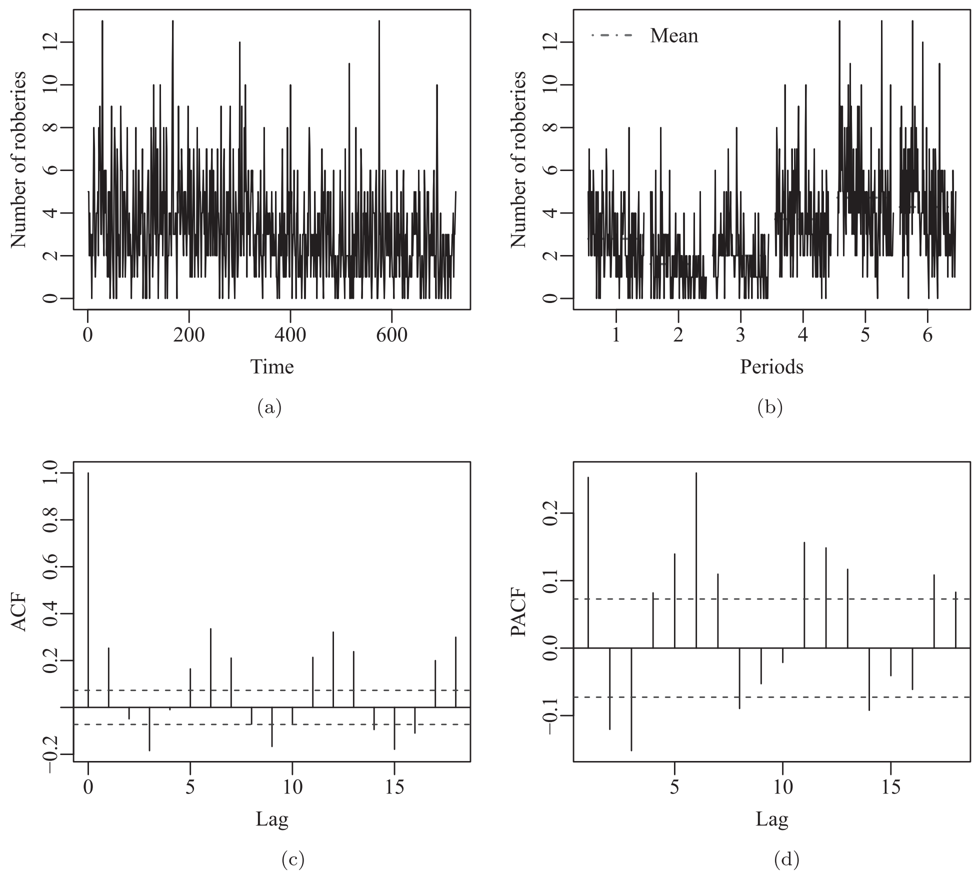

Figure 1. Time series, ACF and PACF plots for number of robberies. (a) Time series. (b) Seasonality. (c) Sample ACF. (d) Sample PACF.

Chicago. We have aggregated the data every 4 hours so as to divide each day into 6 periods, yielding a sample size of $n = 726$.

Some summary statistics of the data are given in Table 2, while Figure 1(a,b) displays the original series and its seasonal component, respectively. It can be observed that the data present seasonal behaviour, with higher values at night time and lower during the day. Figure 1(c,d) presents the sample autocorrelation function (ACF) and partial autocorrelation function (PACF), respectively. These plots indicate the presence of a seasonal dynamic, with a seasonal period $S = 6$, reflecting the effect of the each period of the day in the number of robberies.

The seasonality of this data set motivated our CMP-SARMA model as a candidate model, and for comparison purposes, we also fitted the CMP-ARMA and negative binomial GARMA (NB-GARMA) models to these data. In order to capture the seasonality in the last two models, we introduced the $\sin(2\pi t/6)$ covariate, for $t \in \{1, 2, \ldots, n\}$. In this article, we consider parameterization of the negative binomial distribution given in [30], where the conditional density probability of Y_t is given by

$$Pr(Y_t = y_t \mid \mathcal{F}_{t-1}) = \exp\left\{Y_t \ln(\alpha) + \left(Y_t + \frac{\mu_t}{\alpha}\right)\ln\left(\frac{1}{\alpha+1}\right) + \ln\left(\frac{\Gamma(Y_t + \mu_t/\alpha)}{\Gamma(\mu_t/\alpha)\Gamma(Y_t+1)}\right)\right\},$$

where $\Gamma(\cdot)$ is the gamma function and σ is called the dispersion parameter. The conditional mean and the conditional variance of Y_t 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. The NB-GARMA model was fitted using the *garmaFit* function from *gamlss.util* [31] library in the R software.

Based on the previously mentioned diagnostic measures, the models presented in Table 3 were selected. This table presents the parameter estimates with corresponding standard errors (shown in parentheses), and AIC and SIC values for the selected models. Notice that the CMP-SARMA model presented the smallest AIC and SIC values, suggesting that the proposed model yielded a better fit to the data than the other two models where seasonality is modelled with the deterministic sine covariate. The estimated dispersion parameter indicates overdispersion ($\nu < 1$).

Table 3. Parameter estimates, standard errors (shown in parentheses), and model selection criteria; Chicago robbery data, aggregated by 4-hour periods.

Model		Estimate	AIC	SIC
	$\widehat{\beta}_0$	1.6527(0.1287)		
	$\widehat{\phi}_1$	0.0668(0.0307)		
CMP-SARMA$(1,0) \times (1,1)_{12}$	$\widehat{\Phi}_1$	0.9418(0.0183)	2233.64	2255.16
	$\widehat{\Theta}_1$	0.8422(0.0314)		
	$\widehat{\nu}$	0.7411(0.0603)		
	$\widehat{\beta}_0$	1.1723(0.0448)		
	$\widehat{\beta}_1$	−0.3603(0.0074)		
	$\widehat{\phi}_1$	0.9350(0.0252)		
CMP-ARMA(2, 2)	$\widehat{\phi}_2$	−0.9242(0.0252)	2246.38	2277.50
	$\widehat{\theta}_1$	−0.8727(0.0300)		
	$\widehat{\theta}_2$	0.8968(0.0302)		
	$\widehat{\nu}$	0.7656(0.0609)		
	$\widehat{\beta}_0$	1.1497(0.0276)		
	$\widehat{\beta}_1$	−0.2627(0.1083)		
	$\widehat{\phi}_1$	1.0000(0.0119)		
NB-GARMA(2, 2)	$\widehat{\phi}_2$	−0.9652(0.0166)	2246.97	2277.09
	$\widehat{\theta}_1$	−0.9637(0.0211)		
	$\widehat{\theta}_2$	0.9203(0.0269)		
	$\widehat{\alpha}$	0.0770(0.0223)		

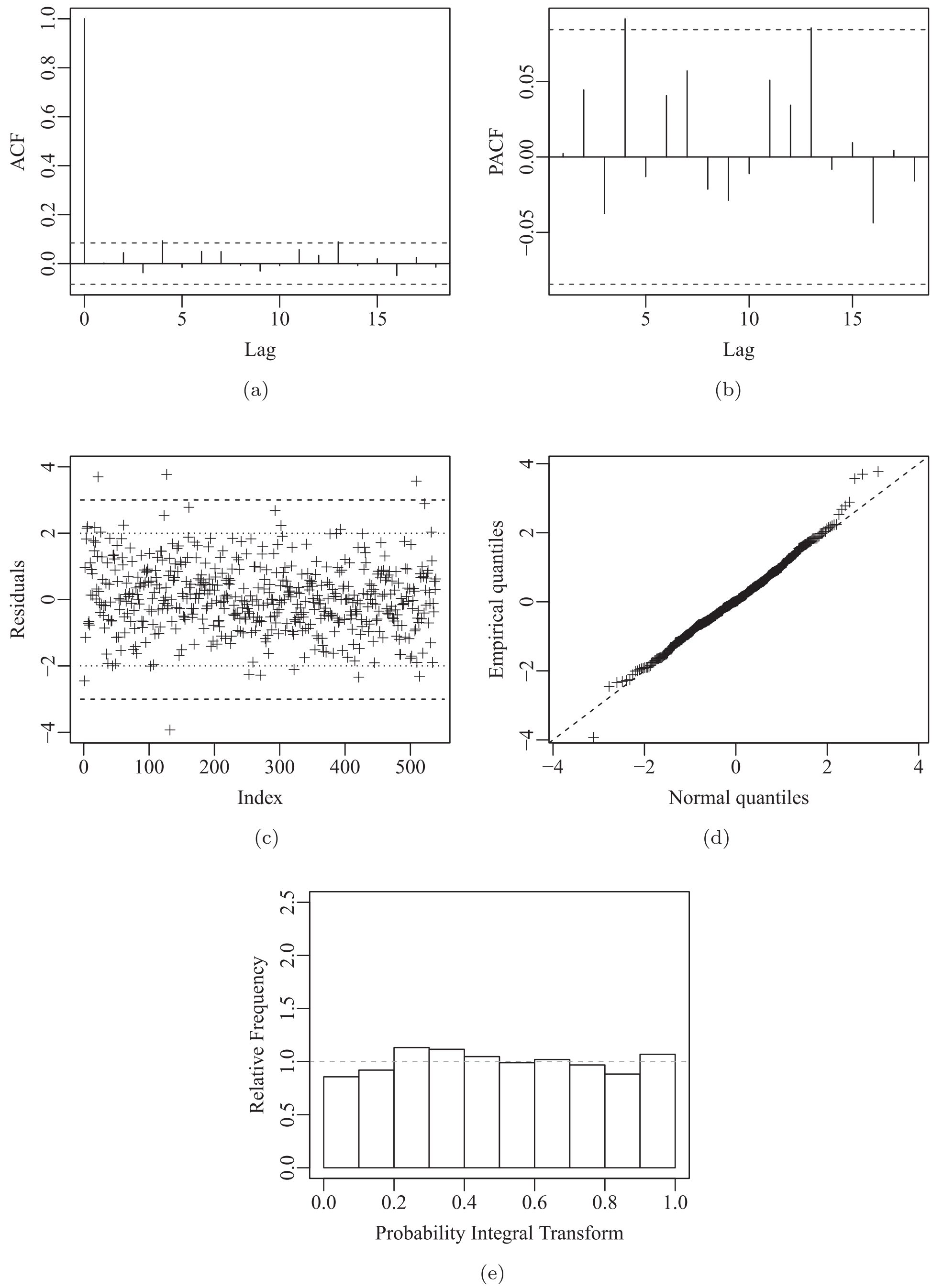

Figure 2. Diagnostic plots for the fitted CMP-SARMA model; Chicago robbery data, aggregated by 4-hour periods. (a) Residual ACF. (b) Residual PACF. (c) Quantile residuals. (d) QQ-plot of the residuals. (e) PIT histogram.

Figure 2 presents a diagnostic analysis of the fitted CMP-SARMA$(1,0) \times (1,1)_{12}$ model. The ACF and PACF of randomized quantile residuals are presented in Figure 2(a ,b), respectively. Notice that there is no indication of significant autocorrelation in the residuals, which is confirmed by the Ljung–Box Q test [32] based on 15 lags. The Ljung–Box test does not reject the null hypothesis with $p-$value $= .2252$. By looking at the residual plot in Figure 2(c), we observe that the residuals are randomly distributed around zero. The plot of the normal against empirical quantiles indicates that the residuals are approximately normally distributed, as shown in Figure 2(d). The uniformity of the PIT in Figure 2(e) (with $J = 10$) indicates that the model was correctly adjusted.

7. Conclusions

The present work proposed the class of Conway–Maxwell–Poisson seasonal autoregressive moving average CMP-SARMA $(p, q) \times (P, Q)_s$ models for time series that can handle both overdispersed and underdispersed counts. This class of models includes seasonal components in which the class of CMP-ARMA models is a special case. We assumed that the conditional distribution of the response variable follows a CMP distribution [17]. We used the conditional maximum likelihood method to estimate the parameters of the proposed model and presented closed-form expressions for the conditional score vector and conditional Fisher information matrix. The asymptotic properties of the ML estimators were established and evaluated based on MC simulations, showing that the estimators are consistent and asymptotically Gaussian. We also discussed practical issues such as diagnostic techniques, hypothesis testing, interval estimation, model selection, and residual analysis. After choosing an initial model based on the Akaike and Bayesian information criteria, a complete residual analysis is necessary to ensure the validity of the assumption that the errors are white noise, that is, they are sequence of serially uncorrelated random variables with zero mean and constant finite variance. Finally, we presented and investigated an empirical application that illustrated the usefulness and applicability of the proposed model.

Note

1. The dataset is available at https://data.cityofchicago.org/.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by CAPES and CNPq.

ORCID

Moizés da Silva Melo http://orcid.org/0000-0001-8966-971X
Airlane Pereira Alencar http://orcid.org/0000-0002-0779-0426

References

[1] Benjamin MA, Rigby RA, Stasinopoulos DM. Generalized autoregressive moving average models. J American Stat Assoc. 2003;98(461):214–223.
[2] Talamantes J, Behseta S, Zender CS. Statistical modeling of valley fever data in Kern County, California. Int J Biometeorology. 2007;51(4):307–313.
[3] Dugas AF, Jalalpour M, Gel Y, et al. Influenza forecasting with Google Flu Trends. PLoS One. 2013;8(2):e56176.
[4] Albarracin OYE, Alencar AP, Lee Ho L. CUSUM chart to monitor autocorrelated counts using negative binomial GARMA model. Stat Meth Medical Res. 2018;27(9):2859–2871.
[5] Albarracin OYE, Alencar AP, Lee Ho L. Effect of neglecting autocorrelation in regression EWMA charts for monitoring count time series. Quality Reliab Eng Int. 2018;34(8):1752–1762.
[6] McCullagh P, Nelder J. Generalized linear models. 2nd ed. Chapman and Hall, London; 1989.
[7] Rocha AV, Cribari-Neto F. Beta autoregressive moving average models. Test. 2009;18(3):529–545.
[8] Bayer FM, Bayer DM, Pumi G. Kumaraswamy autoregressive moving average models for double bounded environmental data. J Hydrology. 2017;555:385–396.
[9] Maior VQ, Cysneiros FJA. Symarma: a new dynamic model for temporal data on conditional symmetric distribution. Stat Papers. 2018;59(1):75–97.
[10] Melo MS, Alencar AP. Conway-Maxwell-Poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data. J Time Ser Anal. 2020. DOI:10.1111/jtsa.12550.
[11] Briët OJ, Amerasinghe PH, Vounatsou P. Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. PLoS One. 2013;8(6):e65761.
[12] Box GE, Jenkins GM, Reinsel GC, et al. Time series analysis: forecasting and control. San Francisco: Holden-Day, John Wiley; 2015.
[13] Tiku ML, Wong WK, Vaughan DC, et al. Time series models in non-normal situations: symmetric innovations. J Time Ser Anal. 2000;21(5):571–596.
[14] Monteiro M, Scotto MG, Pereira I. Integer-valued autoregressive processes with periodic structure. J Stat Planning Inference. 2010;140(6):1529–1541.
[15] Bourguignon M, LP Vasconcellos K, Reisen VA, et al. A Poisson INAR (1) process with a seasonal structure. J Stat Comput Simul. 2016;86(2):373–387.
[16] Bayer FM, Cintra RJ, Cribari-Neto F. Beta seasonal autoregressive moving average models. J Stat Comput Simul. 2018;88(15):2961–2981.
[17] Huang A. Mean-parametrized Conway–Maxwell–Poisson regression models for dispersed counts. Stat Model. 2017;17(6):359–380.
[18] Gay DM. Usage summary for selected optimization routines. Computing Science Technical Report. 1990;153:1–21.
[19] R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/
[20] Andersen EB. Asymptotic properties of conditional maximum-likelihood estimators. J R Stat Soc Ser B (Methodol). 1970;32:283–301.
[21] Pumi G, Valk M, Bisognin C, et al. Beta autoregressive fractionally integrated moving average models. J Stat Planning Inference. 2019;200:196–212.
[22] Pawitan Y. In all likelihood: statistical modelling and inference using likelihood. Oxford University Press, Oxford; 2001.
[23] Kedem B, Fokianos K. Regression models for time series analysis. Vol. 488. Wiley, Hoboken; 2005.
[24] Akaike H. A new look at the statistical model identification. IEEE Trans Automatic Control. 1974;19(6):716–723.
[25] Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6(2):461–464.

[26] Dunn PK, Smyth GK. Randomized quantile residuals. J Comput Graphical Stat. 1996;5(3): 236–244.
[27] Czado C, Gneiting T, Held L. Predictive model assessment for count data. Biometrics. 2009;65(4):1254–1261.
[28] Jung RC, Tremayne A. Useful models for time series of counts or simply wrong ones?. AStA Adv Stat Anal. 2011;95(1):59–91.
[29] Jung RC, McCabe BP, Tremayne AR. Model validation and diagnostics. In: Handbook of discrete-valued time series. Boca Raton (FL): Chapman & Hall/CRC; 2016. p. 189–218.
[30] Evans D. Experimental evidence concerning contagious distributions in ecology. Biometrika. 1953;40(1/2):186–211.
[31] Stasinopoulos MM, Rigby B. Package gamlss.util. page. 2016;9.
[32] Ljung GM, Box GE. On a measure of lack of fit in time series models. Biometrika. 1978;65(2):297–303.