Modelagem de séries temporais de contagem usando a distribuição Conway-Maxwell-Poisson

Moizés da Silva Melo

Tese apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Doutor em Ciências

Programa: Programa de Pós Graduação em Estatística

Orientador: Profa. Dra. Airlane Pereira Alencar

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES e do CNPq

São Paulo, outubro de 2020

Modelagem de séries temporais de contagem usando a distribuição Conway-Maxwell-Poisson

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 23/09/2020. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof[a]. Dr[a]. Airlane Pereira Alencar (IME-USP)
- Prof. Dr. Pedro Alberto Morettin (IME-USP)
- Prof. Dr. Marcelo Bourguignon Pereira (UFRN)
- Prof. Dr. Fábio Mariano Bayer (UFSM)
- Prof[a]. Dr[a]. Linda Lee Ho (EP-USP)

Agradecimentos

Agradeço primeiramente a Deus, que me conduziu até aqui e me deu força em todos os momentos.

À minha orientadora Dra. Airlane Pereira Alencar, pelos ensinamentos passados e por todo apoio oferecido durante a elaboração dessa tese. Sou eternamente grato pelo privilégio de ter sido seu orientando.

Aos meus pais, Francisco Elias Melo (in memoriam) e Josefa Moizés da Silva Melo, que apesar do pouco estudo, sempre me deram suporte e apoio para que eu pudesse estudar e que nunca mediram esforços para que eu pudesse chegar até aqui. Sem eles essa conquista não seria possível.

À minha namorada Laís, por estar sempre comigo durante essa trajetória, compartilhando todos momentos desta grande conquista. Obrigado por estar sempre ao meu lado, pela paciência, pelo amor, pelo carinho.

Aos meus irmãos, Moizaniel da Silva Melo e John Lennon da Silva Melo, que mesmo longe, sempre me apoiaram.

À minha família e à família da Láis. Pelo carinho, amor e incentivo.

Ao professor Dr. Pledson Guedes de Medeiros, meu orientador no mestrado da UFRN, por desempenhar papal fundamental na minha carreira acadêmica.

A todos meus amigos pelos incentivo e apoio, meu muito obrigado.

Aos funcionários do IME-USP, em especial a Regiane, secretária do programa.

Ao Programa de Pós Graduação em Estatística do Instituto de Matemática e Estatística da Universidade de São Paulo pela oportunidade concedida

À Banca Examinadora pelas sugestões.

À CAPES e CNPq pelo apoio financeiro concedido.

Resumo

MELO, M. S. **Modelagem de séries temporais de contagem usando a distribuição Conway-Maxwell Poisson**. 2020. 87 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

Recentemente, modelos para dados de séries temporais que não satisfazem a suposição de normalidade vêm sendo propostos a fim de possibilitar melhores ajustes a dados reais. O presente trabalho tem por objetivo desenvolver modelos de séries temporais e propor gráficos de controle para dados de contagem com subdispersão, equidispersão e sobredispersão, baseado na distribuição Conway-Maxwell-Poisson. A distribuição Conway-Maxwell-Poisson é bastante flexível e tem como casos particulares as distribuições Poisson e geométrica e como caso limite a distribuição binomial. Sua principal vantagem quando comparada a outras distribuições discretas é que permite acomodar adequadamente a subdispersão e sobredispersão encontrada frequentemente na análise de conjuntos de dados reais. Este trabalho é composto por 3 artigos. Nos dois primeiros artigos, propomos dois novos modelos de séries temporais de contagens, intitulados Modelo Conway-Maxwell-Poisson Autorregressivo de Médias Móveis para dados de contagem subdispersos, equidispersos e sobredispersos; e Modelo Conway-Maxwell-Poisson Sazonal Autorregressivo de Médias Móveis, respectivamente. Apresentamos os estimadores de máxima verossimilhança condicional, teste de hipóteses e análise de diagnóstica para os modelos propostos. Realizamos estudos de simulação para verificar as propriedades de amostras finitas dos estimadores. Os resultados numéricos mostram que os estimadores dos dois modelos propostos possuem boas propriedades assintóticas, à medida que o tamanho amostral aumenta o viés e o erro quadrático médio de todos os estimadores diminuem. Também fornecemos expressões de forma fechada para o vetor escore condicional e a matriz de informações de Fisher condicional. Finalmente, ilustramos a utilidade dos modelos propostos, explorando aplicações empíricas. No último artigo, propomos um novo gráfico de controle com memória para monitorar dados de contagem autocorrelacionados, no qual uma média progressiva é usada como a estatística de plotagem. O novo gráfico de controle é baseado nos resíduos quantílicos aleatorizados obtidos pelo ajuste de um modelo Conway-Maxwell Poisson Autorregressivo de Média Móvel. Um estudo de simulação é realizado para avaliar o desempenho do gráfico de controle proposto. Os resultados mostram que a proposta atual apresenta bom desempenho para detectar pequenas, moderadas e grandes mudanças na média do processo. Além disso, quando comparado com os gráficos de controle do tipo EWMA e Shewhart, o novo gráfico apresentou desempenho melhor em termo do número médio de amostras até o sinal.

Palavras-chave: Séries temporais, Distribuição Conway-Maxwell-Poisson, Gráfico de controle, Dados de contagem, Sobredispersão, Subdispersão.

Abstract

MELO, M. S. **Count time series modeling using the Conway-Maxwell Poisson distribution**. 2020. 87 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

Recently, models for time series data that do not satisfy the assumption of normality have been proposed to be able to fit real data better. The present work aims to develop regression models and propose control charts for time series of counts with subdispersion, equidispersion and overdispersion, based on the Conway-Maxwell-Poisson distribution. The Conway-Maxwell-Poisson distribution is very flexible and it has the Poisson and geometric distributions as special cases, and the Bernoulli distribution as a limiting case. Its main advantage when compared to other discrete distributions is that it can adequately accommodate the subdispersion and overdispersion found frequently in the analysis of real data sets. This thesis is composed by 3 articles. In the first two articles, we propose two new models for time series of counts, namely Conway-Maxwell-Poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data; and Conway-Maxwell-Poisson seasonal autoregressive moving average model, respectively. We presented and discussed the conditional maximum likelihood estimators, hypothesis testing, and some diagnostic tools for the proposed models. We conduct a Monte Carlo simulation to evaluate of the finite sample performance of the proposed estimators. The numerical results show good asymptotic properties of the estimadors, as the sample size increases, the bias and mean square error of all estimators decreases. We also provide closed forms for the conditional score vector and the conditional Fisher information matrix. Finally, we illustrate the usefulness of the proposed models, exploring empirical applications. In the last article, we propose a new memory-type control chart by monitoring autocorrelated count data, in which a progressive mean is used as the plotting statistic. The new control chart is based on randomized quantile residuals obtained from a fitted Conway-Maxwell Poisson autoregressive moving average model. A simulation study is carried out to evaluate the performance of the proposed control chart. The results show that the current proposal performs well to detect small, moderate, and large shifts in the process mean. In addition, when compared to the control charts of type EWMA and Shewhart, the new chart showed better performance in terms of average run lengths(ARL).

Keywords: Time series, Conway-Maxwell-Moisson distribution, Control chart, Count data, Overdispersion, Underdispersion.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The most well-known and frequently used time series models in the literature are based on he Gaussianity assumption (Box *et al.*, 2015). However, the assumption of normality is not always satisfied in practice. As a consequence, in recent years, there has been increasing interest in the development and analysis of non-Gaussian models for time. Benjamin *et al.* (2003) introduced the class of generalized linear autoregressive moving average (GARMA) models, which is an extension of the standard ARMA models for situations where the conditional distribution of the dependent variable given the past of the process belongs to the exponential family, such as the Poisson and binomial negative distributions. These class of models are widely used to model discrete response time series (Albarracin *et al.*, 2018b). In practice, however, Poisson and negative binomial distributions are not suitable for modeling underdispersed count data, where the variance is smaller than the mean. There are some distributions that contain the Poisson distribution as special case and can to model overdispersed and underdispersion count data. For example, the generalized Poisson, the gamma count, the double Poisson, and the Conway-Maxwell Poisson (CMP or COM-Poisson) distributions (Sellers and Morris, 2017).

Recent works on time series modeling has used the CMP distribution to model count data with equidispersion, underdispersion and overdispersion. For example, Zhu (2012) introduced a CMP integer generalized autoregressive conditionally heteroscedastic (INGARCH) model; Mamode Khan *et al.* (2018) proposed an observation-driven longitudinal integer-valued moving average model of order 1 (INMA(1)) with CMP innovations; Sunecher *et al.* (2018) introduced a first-order bivariate integer-valued moving average (BINMA(1)) models driven by CMP innovations; among others. The CMP distribution is very flexible and generalizes several classical distributions, such as the geometric and Poisson distributions as special cases, and the Bernoulli distribution as a limiting case. This distribution's flexibility and special properties encourage its empirical use in a wide range of applications (Sellers *et al.*, 2012).

The CMP distribution has also attracted considerable attention in the quality control area. Several control charts have been proposed in the literature to monitor count data using the CMP distribution. Sellers (2012) proposed a Shewhart-type control chart based on the CMP distribution. Saghir and Lin (2014a) introduced a Shewhart-type multivariate control chart is constructed to monitor such kind of data based on the CMP distribution. Saghir and Lin (2014b) proposed cumulative sum (CUSUM) charts for monitoring the CMP processes. Aslam *et al.* (2016) introduced an attribute control chart for the CMP distributed non-conformities using multiple dependent states sampling based on the modified exponentially weighted moving average (EWMA) statistic.

Aslam *et al.* (2017) developed a control chart for the CMP distribution using the EWMA statistic. Aslam *et al.* (2018) proposed a hybrid EWMA control chart for count data assuming CMP distribution. Alevizakos and Koukouvinos (2019a) presented a double EWMA control chart with steady-state control limits to monitor CMP attributes. Alevizakos and Koukouvinos (2019b) introduced a control chart based on the progressive mean (PM) statistic for monitoring CMP distributed data. Rao *et al.* (2020) developed a mixed EWMA-CUSUM control chart for monitoring moderate and/or small shift in the process following the CMP distribution.

The CMP distribution, originally developed by Conway and Maxwell (1962) as a model for queuing systems with state dependent service times, had its statistical properties, as well as the methods to estimate its parameters established by Shmueli *et al.* (2005). Due to its flexibility and nice special properties, the CMP distribution has been used in different areas of application such as transport, linguistic, marketing and many others (Lord *et al.*, 2008, 2012; Sellers *et al.*, 2012). In the original formulation, the CMP 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 considered the dispersion parameter such that $\nu > 1$ represents underdispersion and $0 \leq \nu < 1$ overdispersion. When $\nu = 1$, the distribution coincides with the Poisson distribution.

Although the CMP distribution is very flexible for modeling count data, its moments cannot be expressed in terms of the two parameters that index the distribution, λ and ν. The inability to model the mean directly limited the use of the CMP distribution in practice, especially in regression analysis, where it is typically more useful to model the mean of the distribution (Huang, 2017). Recently, in order to obtain a regression structure for the mean of the CMP distribution, and thus facilitate the interpretation of the regression models, Huang (2017) proposed a mean-parametrization of the CMP distribution. In the proposed model by Huang (2017), the regression coefficients are directly associated with the mean of the response variable, as in the class of the generalized linear models (GLMs) introduced by McCullagh (1984). In addition, in this parameterization the mean and dispersion parameters are orthogonal. As a consequence, the maximum likelihood estimators for the mean and dispersion parameter are asymptotically independent.

The reparameterized CMP distribution (Huang, 2017), denoted by CMP_μ, is given by the probability mass function

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

where $\lambda(\mu, \nu)$ is obtained by solving the following non-linear equation

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

The purpose of this thesis is to propose two dynamic regression models for counting data with overdispersion and underdispersion based on the CMP distribution with time-varying conditional mean depending on covariates and lagged observations. Following the approach presented in the GARMA model, in the first model, the conditional mean of the dependent variable is modeled

through a autoregressive moving average structure. This model can be viewed with a combination of the ARMA model and the generalized linear model (GLM) (McCullagh, 1984). In the second article, we extended the class of models proposed in the first article including seasonality components. In addition, we proposed a new memory-type control chart, in which a progressive mean is used as the plotting statistic for monitoring autocorrelated count data. The development of the proposed control chart is based on the randomized quantile residuals obtained from a fitted model.

1.1 Contributions

The main contributions of this work are the following:

- To propose dynamic regression models based on the Conway-Maxwell-Poisson distribution for modeling underdispersed, equidispersed, and overdispersed time series count data.

- To evaluate the asymptotic properties of conditional maximum likelihood estimators of the models, as well as to provide closed-form expressions for the conditional score vector and conditional Fisher information matrix.

- To propose memory-type control chart based on the progressive mean statistic to monitor autocorrelated count time series. The new control chart is based on the randomized quantile residuals obtained from a fitted Conway-Maxwell-Poisson autoregressive moving average model.

1.2 Thesis organization

The thesis is composed by 3 articles. The first article, Conway-Maxwell-Poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data, is introduced in Chapter 2. The second article, Conway-Maxwell-Poisson seasonal autoregressive moving average model, is presented in Chapter 3. And the last article, Progressive Mean Control Chart for Monitoring Count Time Series is developed in Chapter 4. Each chapter in the thesis can be read independently.

Chapter 2

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

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; Overdispersion; Underdispersion

2.1 Introduction

Models for time series of counts have received considerable and growing attention in recent decades. These series are commonly observed in real applications such as economics, medicine, and epidemiology (Franke and Seligmann, 1993; Freeland and McCabe, 2004; 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 (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 dispersion 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 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 article, the mean of the conditional distribution is directly modeled, making the model parameters easily interpretable.

The present article introduces a dynamic regression model for time series following a CMP distribution. 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 (Bayer *et al.*, 2017; Fokianos and Kedem, 2004; Li, 1991, 1994; Rocha and Cribari-Neto, 2009).

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 parametrization 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 article is organized as follows. Section 2.2 reviews the CMP distribution and its reparametrization, where the mean of the distribution is rewritten as a function of the original parameters. Section

2.3 introduces the proposed model. In Section 2.4, 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 2.5 discusses some diagnostic measures and forecasting. In Section 2.6, we present the results of the Monte Carlo simulation study. Section 2.7 illustrates the flexibility of the proposed model through two empirical applications. Finally, some conclusions are given in Section 2.8. Details on the derivation of the conditional score vector and conditional Fisher information matrix are presented in the Appendices.

2.2 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

$$E(Y) \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu} \text{ and } V(Y) \approx \frac{\lambda^{1/\nu}}{\nu} \tag{2.1}$$

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

Sellers and Shmueli (2010) proposed the CMP regression model using the original parametrization $\log(\lambda_i) = x_i^T\beta$, where the mean is modeled through the approximation in (2.1). Note that this approximation is accurate for $\nu \leq 1$ or $\lambda > 10^\nu$, and the mean is indirectly modeled. 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 GLMs. This reparametrization makes the model simpler and easily interpretable.

The reparametrized CMP distribution (Huang, 2017), denoted by CMP_μ, is given by the probability mass function

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

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

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

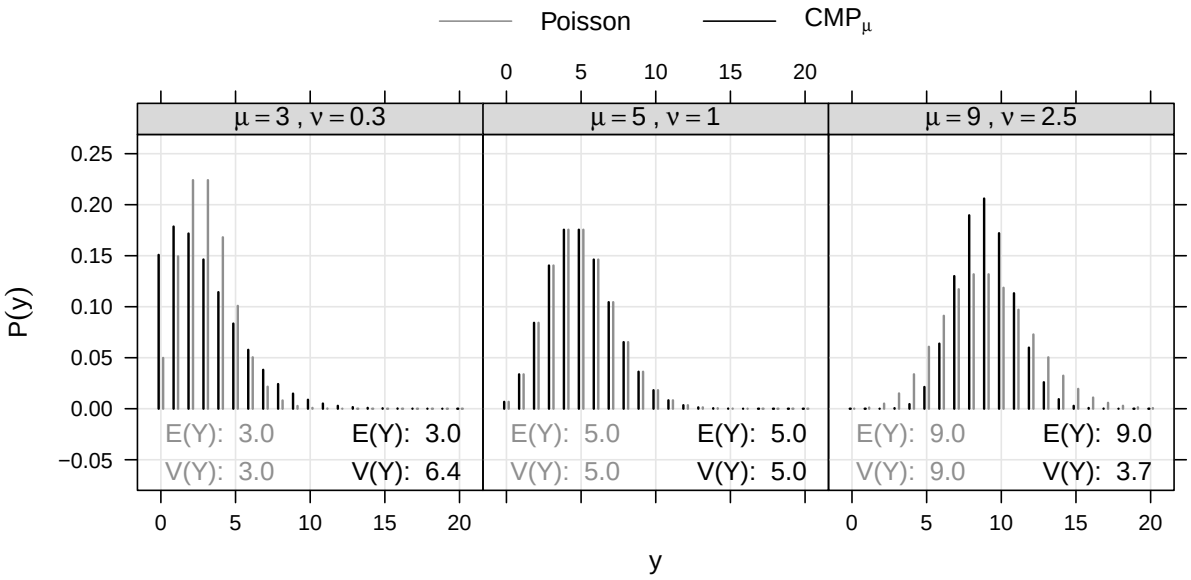


Figure 2.1: *Shapes of the Poisson and CMP$_\mu$ probability mass functions for different values of μ and ν*

.

In Figure 2.1, 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$_\mu$ distribution belongs to the one parameter exponential family.

2.3 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 (1).

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 $\mathcal{F}_{t-1} = \{Y_{t-1}, \ldots, Y_1\}$, 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

$$Pr(Y_t = y_t \mid \mathcal{F}_{t-1}, \mu_t, \nu) = \frac{\lambda(\mu_t, \nu)^{y_t}}{(y_t!)^\nu Z(\lambda(\mu_t, \nu), \nu)}, \quad y = 0, 1, 2, \ldots, \tag{2.2}$$

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

$$0 = \sum_{s=0}^{\infty} (s - \mu_t) \frac{\lambda^s}{(s!)^\nu}, \tag{2.3}$$

and $Z(\lambda(\mu_t, \nu), \nu) = \sum_{s=0}^{\infty} \lambda(\mu_t, \nu)^s / (s!)^\nu$ is a normalization function, 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

$$\eta_t = g(\mu_t) = \alpha + \boldsymbol{x}_t^\top \boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j [g(y_{t-j}) - \boldsymbol{x}_{t-j}^\top \boldsymbol{\beta}] + \sum_{j=1}^{q} \theta_j r_{t-j}, \tag{2.4}$$

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 article, we consider the errors measured on the predictor scale $r_t = g(y_t) - g(\mu_t)$ as in Bayer *et al.* (2017), although other the moving average error terms can be used Benjamin *et al.* (2003); Rocha and Cribari-Neto (2009).

The proposed CMP-ARMA(p, q) model is defined by (2.2) and (2.4). 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. To allow the use of the logarithm link function with modeling count series containing observations equal to zero, we replace y_{t-j} in (2.4) 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 (2.4) 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).

2.4 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

$$\ell(\boldsymbol{\gamma}) = \sum_{t=m+1}^{n} \log Pr(y_t|\mathcal{F}_{t-1}, \mu_t, \nu_t) = \sum_{t=m}^{n} \ell_t(\mu_t, \nu), \tag{2.5}$$

where

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

2.4.1 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

$$\begin{aligned}
U_\alpha(\boldsymbol{\gamma}) &= \boldsymbol{\delta}^\top \boldsymbol{T}\boldsymbol{V}(\boldsymbol{y} - \boldsymbol{\mu}), \\
\boldsymbol{U}_\beta(\boldsymbol{\gamma}) &= \boldsymbol{M}^\top \boldsymbol{T}\boldsymbol{V}(\boldsymbol{y} - \boldsymbol{\mu}), \\
\boldsymbol{U}_\phi(\boldsymbol{\gamma}) &= \boldsymbol{P}^\top \boldsymbol{T}\boldsymbol{V}(\boldsymbol{y} - \boldsymbol{\mu}), \\
\boldsymbol{U}_\theta(\boldsymbol{\gamma}) &= \boldsymbol{Q}^\top \boldsymbol{T}\boldsymbol{V}(\boldsymbol{y} - \boldsymbol{\mu}), \\
U_\nu(\boldsymbol{\gamma}) &= \sum_{t=m+1}^{n} E_{\mu_t,\nu}\left[\log(y_t!)(\mu_t - y_t)\right] \frac{(y_t - \mu_t)}{V(\mu_t, \nu)} - \left[\log(y_t!) - E_{\mu,\nu} \log(y_t!)\right],
\end{aligned} \tag{2.6}$$

where $\boldsymbol{y} = (y_{m+1}, \ldots, y_n)^\top$, $\boldsymbol{\mu} = (\mu_{m+1}, \ldots, \mu_n)^\top$, $\boldsymbol{T} = \text{diag}\{1/g'(\mu_{m+1}), \ldots, 1/g'(\mu_n)\}$, $\boldsymbol{\delta} = \left(\frac{\partial \eta_{m+1}}{\partial \alpha}, \ldots, \frac{\partial \eta_n}{\partial \alpha}\right)^\top$, $\boldsymbol{V} = \text{diag}\{1/V(\mu_{m+1}, \nu), \ldots, 1/V(\mu_n, \nu)\}$, $\boldsymbol{M}$ 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 (2.4) is defined by $r_t = g(y_t) - g(\mu_t)$, then

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

(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 $\boldsymbol{U}(\boldsymbol{\gamma}) = \mathbf{0}$, where $\mathbf{0}$ 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` (R Core Team, 2019), which is a reverse-communication trust-region quasi-Newton method from the **Port** library (Gay, 1990).

2.4.2 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 (2.5). Let $W = \text{diag}\{w_1, \ldots, w_n\}$, where

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

The conditional Fisher information matrix is given by

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

(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,

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\phi}} \\ \widehat{\boldsymbol{\theta}} \\ \widehat{\nu} \end{pmatrix} \sim N_{r+p+q+2} \left(\begin{pmatrix} \alpha \\ \boldsymbol{\beta} \\ \boldsymbol{\phi} \\ \boldsymbol{\theta} \\ \nu \end{pmatrix}, \boldsymbol{K}^{-1} \right), \tag{2.7}$$

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.

2.4.3 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 (2.7), 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, vs. the alternative hypothesis $\mathcal{H}_1 : \gamma_i \neq \gamma_i^0$. A convenient statistic to test individual parameters is the so-called z statistic (Pawitan, 2001) given by

$$Z = \frac{\widehat{\gamma}_i - \gamma_i}{\sqrt{k^{ii}}}. \tag{2.8}$$

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 (2.8) can be used to test equidispersion in the data. We considered the following hypotheses

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

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.

2.5 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.

2.5.1 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).

2.5.2 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

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

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

2.5.3 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, that is, 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.

2.5.4 Probability Integral Transform 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.*, 2016). 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 (2.2).

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

$$f_j = \overline{F}\left(\frac{j}{J}\right) - \overline{F}\left(\frac{j-1}{J}\right).$$

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

2.5.5 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 (2.4),

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

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 prediction intervals for the forecasts using quantiles.

2.6 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.

2.6.1 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 on 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 2.1 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.

2.6.2 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 $\Phi(B) = 1 - \phi_1 B - \cdots, \phi_p B^p$, where B is a backshift operator ($B^d y_t = y_{t-d}$). When the link function is different from identity, the parameter restrictions to ensure stationarity appear to be analytical intractable.

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 non-stationarity in each parameter combination. We found no evidence of non-stationarity for any of the parameter combinations considered.

2.7 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

Table 2.1: *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
	Mean	1.6873	0.5052	0.4350	0.3189	0.5757
$n = 50$	RB(%)	12.4923	1.0419	−12.9926	6.3083	15.1501
	MSE	0.3605	0.0129	0.0404	0.0489	0.0463
	Mean	1.5997	0.5012	0.4655	0.3099	0.5350
$n = 100$	RB(%)	6.6471	0.2445	−6.8854	3.3208	7.0074
	MSE	0.1457	0.0062	0.0162	0.0190	0.0156
	Mean	1.5518	0.5003	0.4822	0.3056	0.5169
$n = 200$	RB(%)	3.4570	0.0491	−3.5533	1.8748	3.3947
	MSE	0.0634	0.0031	0.0070	0.0084	0.0062
	Mean	1.5269	0.5003	0.4908	0.3025	0.5082
$n = 400$	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
	Mean	1.6937	0.5040	0.4336	0.3208	1.1494
$n = 50$	RB(%)	12.9184	0.8036	−13.2679	6.9661	14.9459
	MSE	0.3720	0.0090	0.0416	0.0507	0.0918
	Mean	1.6029	0.5009	0.4648	0.3111	1.0689
$n = 100$	RB(%)	6.8606	0.1979	−7.0274	3.6943	6.8913
	MSE	0.1535	0.0043	0.0170	0.0200	0.0314
	Mean	1.5520	0.5009	0.4823	0.3059	1.0326
$n = 200$	RB(%)	3.4692	0.1979	−3.5334	1.9905	3.2641
	MSE	0.0664	0.0021	0.0073	0.0089	0.0125
	Mean	1.5266	0.5007	0.4909	0.3031	1.0157
$n = 400$	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
	Mean	1.7001	0.5021	0.4324	0.3214	2.2982
$n = 50$	RB(%)	13.3400	0.4378	−13.5150	7.1332	14.9109
	MSE	0.3845	0.0049	0.0428	0.0536	0.3661
	Mean	1.6079	0.5006	0.4636	0.3119	2.1367
$n = 100$	RB(%)	7.1985	0.1298	−7.2750	3.9884	6.8370
	MSE	0.1617	0.0021	0.0179	0.0209	0.1232
	Mean	1.5560	0.5003	0.4811	0.3070	2.0658
$n = 200$	RB(%)	3.7369	0.0556	−3.7651	2.3506	3.2930
	MSE	0.0698	0.0010	0.0077	0.0093	0.0500
	Mean	1.5282	0.5001	0.4905	0.3034	2.0309
$n = 400$	RB(%)	1.8850	0.0398	−1.8974	1.1381	1.5470
	MSE	0.0350	0.0005	0.0038	0.0044	0.0226

Binomial GARMA models, considering the parametrization of the Negative Binomial distribution used in Evans (1953). Using this parametrization, 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.

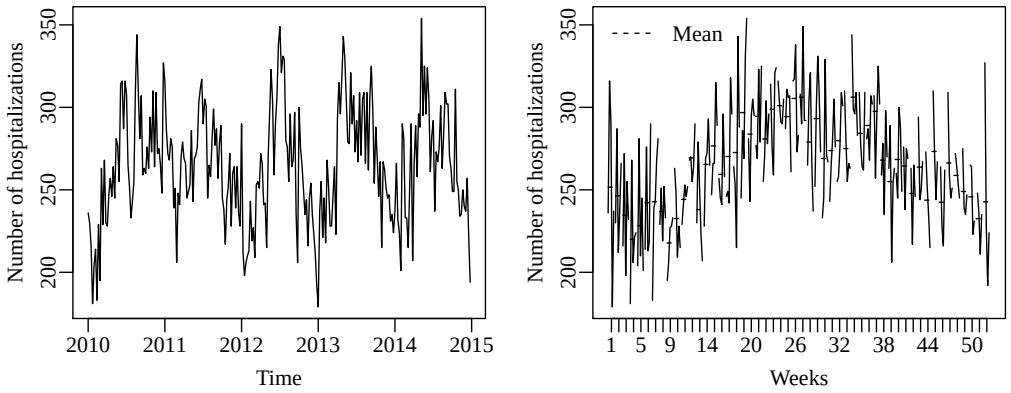
2.7.1 Overdispersion Data: Weekly Number of Hospitalizations

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/`).

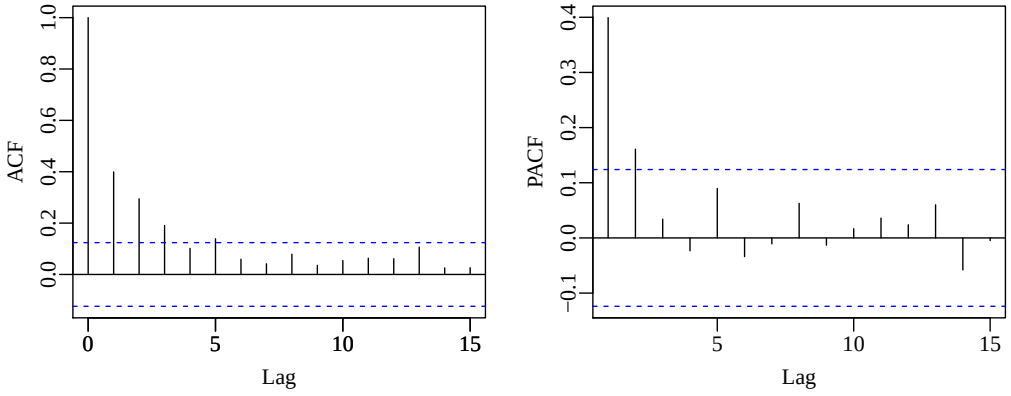
The empirical mean and variance of the data are 264.61 and 1201.60, respectively, indicating that the data are overdispersed. The original series and its seasonal component are plotted in Figures 2.2(a) and 2.2(b), respectively. The data present seasonal behavior since the mean number of hospitalizations 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 2.3(a) and 2.3(b), respectively, indicate a second-order autoregressive autocorrelation structure. Hence, we fitted the proposed model to 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 2.2 shows the parameter estimates and corresponding standard errors (SE), z statistics, p-values and information criteria. For comparison purposes, Table 2.2 also shows the fitted Poisson and Negative Binomial GARMA$(2, 0)$ models, which were the best Poisson and Negative Binomial GARMA models for this data set, that is, the models with the lowest AIC and/or BIC. 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 $\mathcal{H}_0 : \nu = 1$ vs. $\mathcal{H}_1 : \nu \neq 1$ and rejected the null



(a) Observed values (b) Seasonality

Figure 2.2: *(a) Observed number of hospitalizations and (b) seasonal component present in the data.*



(a) Residual ACF (b) Residual PACF

Figure 2.3: *(a) ACF and (b) PACF of randomized quantile residuals for the CMP-GLM model.*

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 2.4. Figures 2.4(a) and 2.4(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 2.4(c) seems to be oscillating around zero with constant variance. Figure 2.4(d) presents the empirical and normal quantiles. The analysis of these two plots indicates that the residuals are approximately normally distributed. Figure 2.4(e) shows the non-randomized PIT histogram of the fitted model with $J = 10$. The uniformity of the PIT (Figure 2.4(e)) suggests that the CMP-ARMA$(2, 0)$ is a suitable model to the data, and Figure 2.6(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 2.5 presents diagnostic plots for the Poisson GARMA model. Although the ACF (Figure 2.5(a)) and PACF (Figure 2.5(b)) of randomized quantile residuals indicate that residual autocorrelations are not significant, the sequence of residuals (Figure 2.5(c)) and normal probability (Figure 2.5(d))

Table 2.2: *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 GARMA	$\widehat{\beta}_0 = 5.5818$	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		

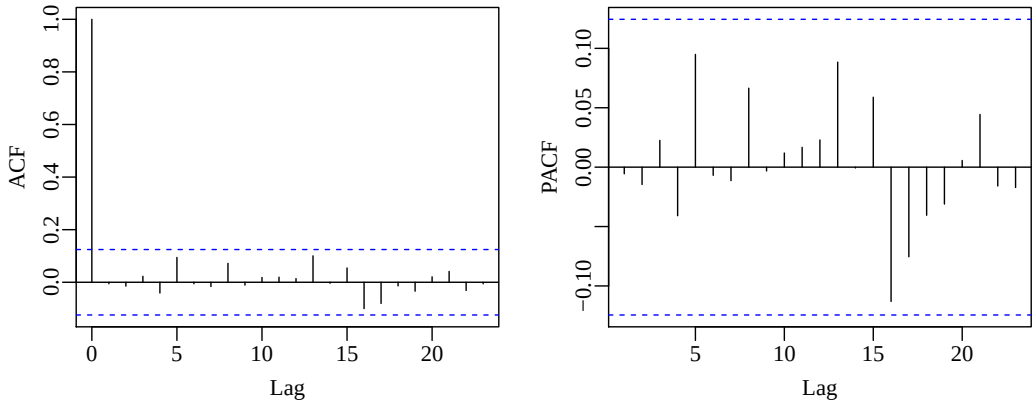
plots indicate that the residuals are non-normally distributed. Furthermore, the PIT histogram (Figure 2.5(e)) and the normal probability plots suggest that the Poisson GARMA model is unable to capture overdispersion in the data.

Finally, the out-of-sample forecasts using the fitted CMP-ARMA model are presented in Figure 2.6(b) along with the observed values for comparison reason. The mean absolute percentage error between the observed data (y_{n+h}) and out-of-sample forecasts $(\widehat{y}_{n+h})$, for $h = 1, \ldots, 10$, is 5.08%.

2.7.2 Underdispersed Data: Pedestrian Counts

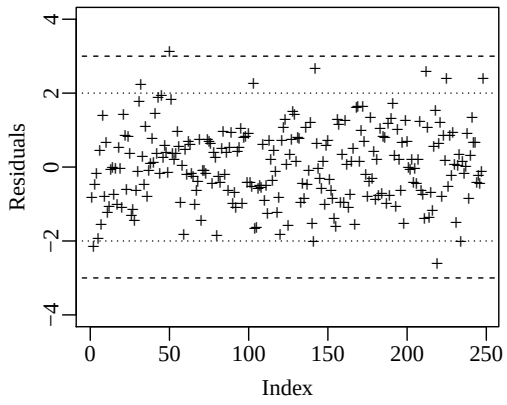
We analyzed the second data set to illustrate the flexibility of the proposed model for underdispersed data. The data set consists of 505 counts of the number of pedestrians traversing a city block observed at 5-second intervals. These data were originally presented by Fürth (1918) and later analyzed 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 2.7 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 2.8(a) and 2.8(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.

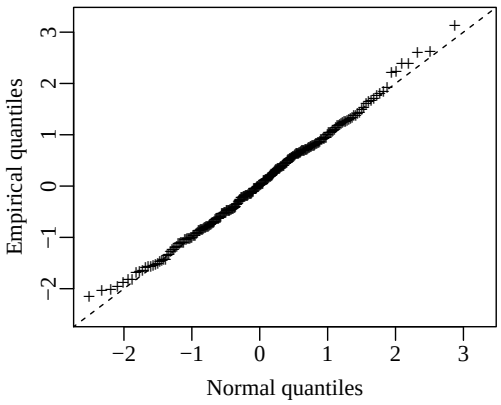


(a) Residual ACF

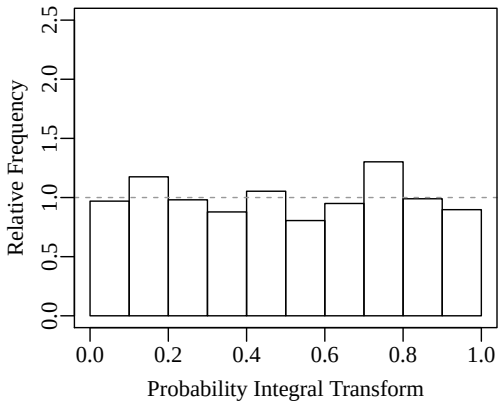
(b) Residual PACF



(c) Quantile residuals



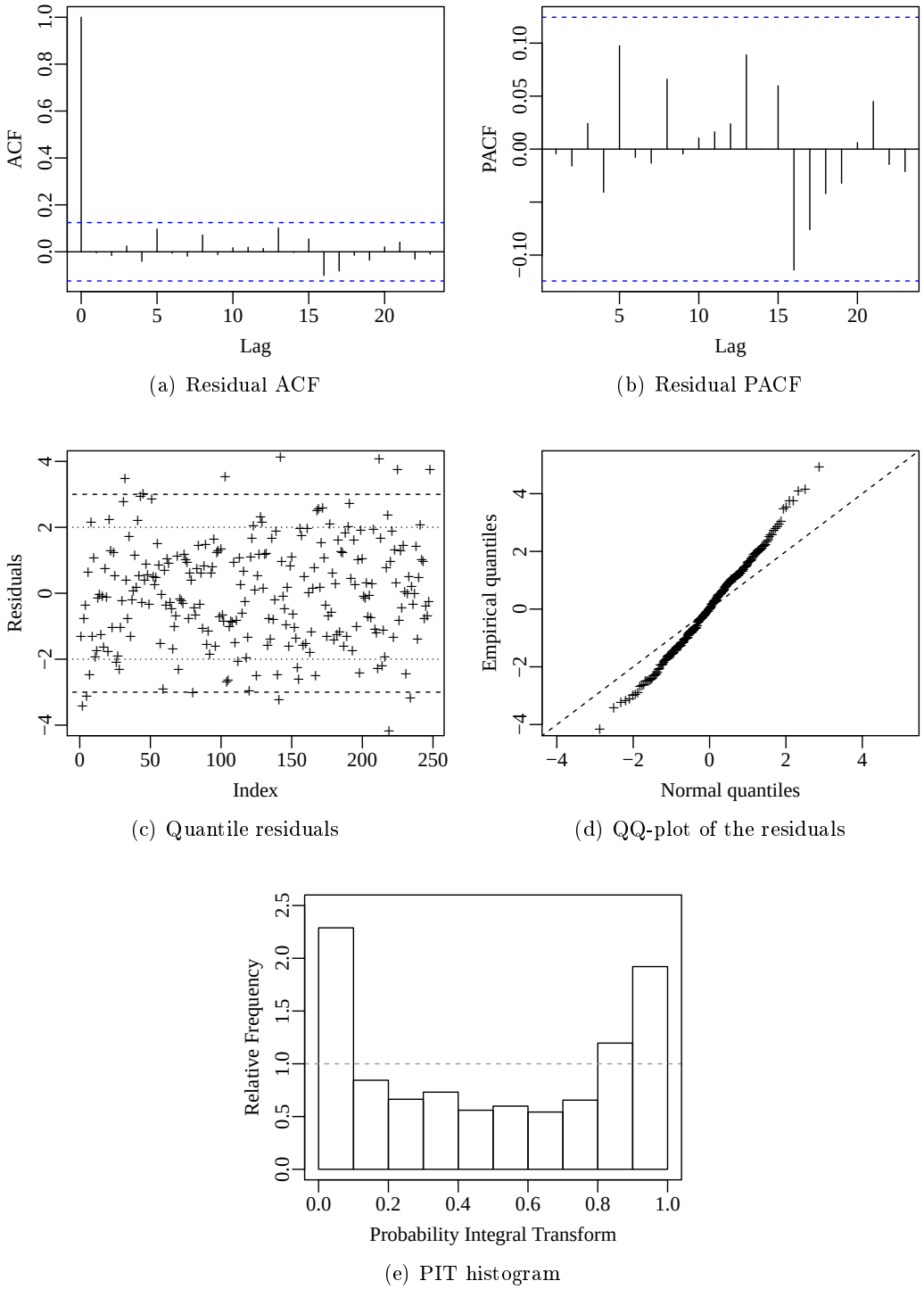
(d) QQ-plot of the residuals



(e) PIT histogram

Figure 2.4: *Diagnostic plots for the fitted CMP-ARMA model; weekly number of hospitalizations data.*

Table 2.3 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 $\mathcal{H}_0 : \nu = 1$ vs. $\mathcal{H}_1 : \nu \neq 1$ 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



(a) Residual ACF

(b) Residual PACF

(c) Quantile residuals

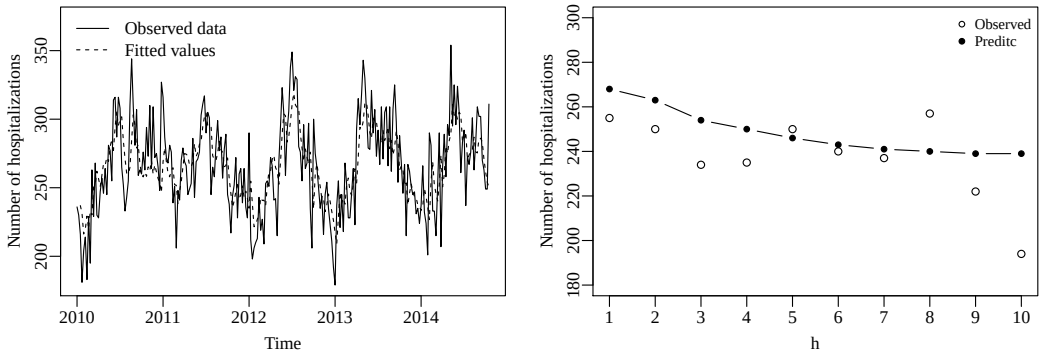
(d) QQ-plot of the residuals

(e) PIT histogram

Figure 2.5: *Diagnostic plots for the fitted Poisson-GARMA model; weekly number of hospitalizations data.*

Negative Binomial GARMA models.

Figure 2.9 shows diagnostic plots for the CMP-ARMA model. The ACF and PACF of randomized quantile residuals shown in Figures 2.9(a) and 2.9(b), respectively, along with Box-Ljung statistic based on 15 lags, confirm the assumption that there is no significant residual autocorrelation



(a) Fitted values (b) Forecasts

Figure 2.6: *(a) Fitted values and (b) forecasts; weekly number of hospitalizations data.*

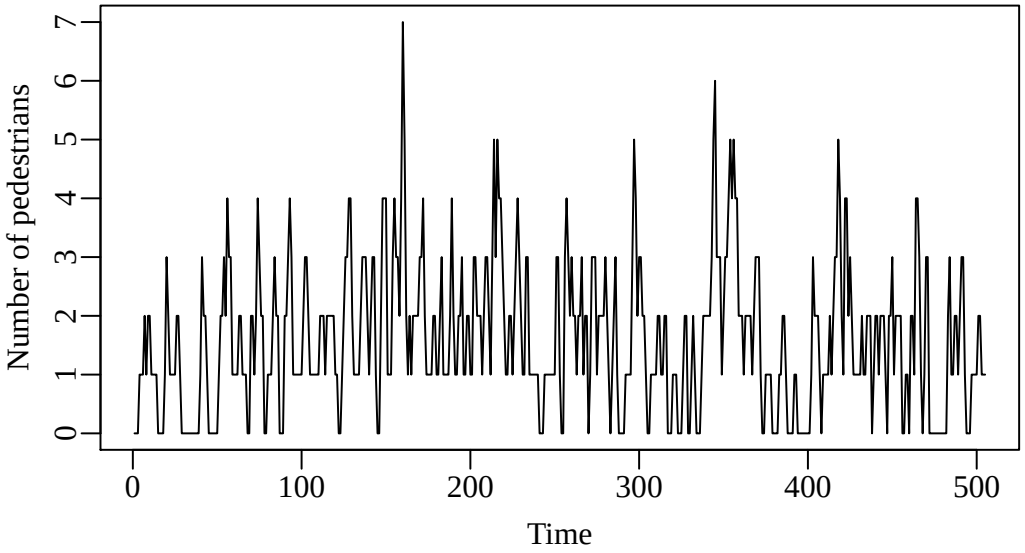
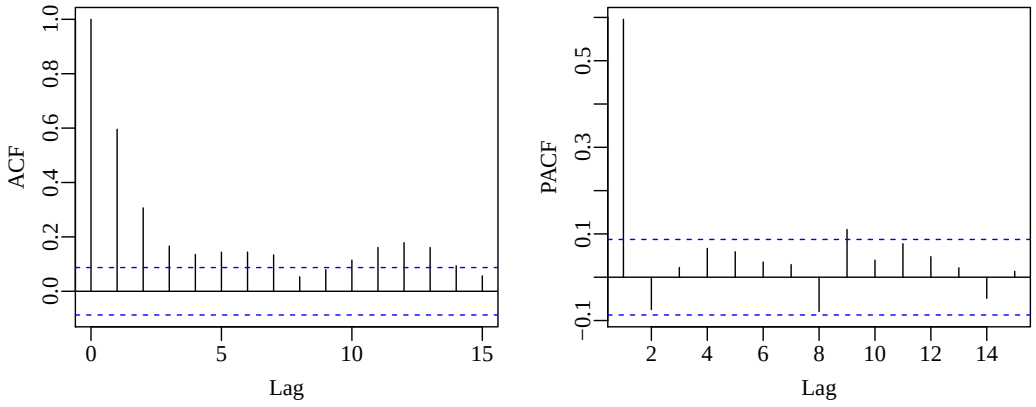


Figure 2.7: *Number of pedestrians on a city block observed every* 5 *seconds*

(p-value $= 0.146$). Figures 2.9(c) and 2.9(d) show the residual and normal probability plots, respectively. Both plots indicate that the residuals are normally distributed. Figure 2.9(e) displays the PIT histogram, indicating that the fitted CMP-ARMA$(1,1)$ model is correctly specified. Figures 2.10 and 2.11 present diagnostic plots for the Poisson and Negative GARMA models, respectively. The ACF and PACF of randomized quantile residuals (Figures 2.10(a), 2.10(b), 2.11(a), and 2.11(b)) indicate that residual autocorrelations are not significant, and Figures 2.10(c) and 2.11(c) show that the residuals are randomly distributed around zero for the two models. However, the normal probability and PIT histogram plots (Figures 2.10(d), 2.10(e), 2.11(d), and 2.11(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 2.12 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

Figure 2.8: *(a) ACF and (b) PACF of randomized quantile residuals for the CMP-GLM model.*

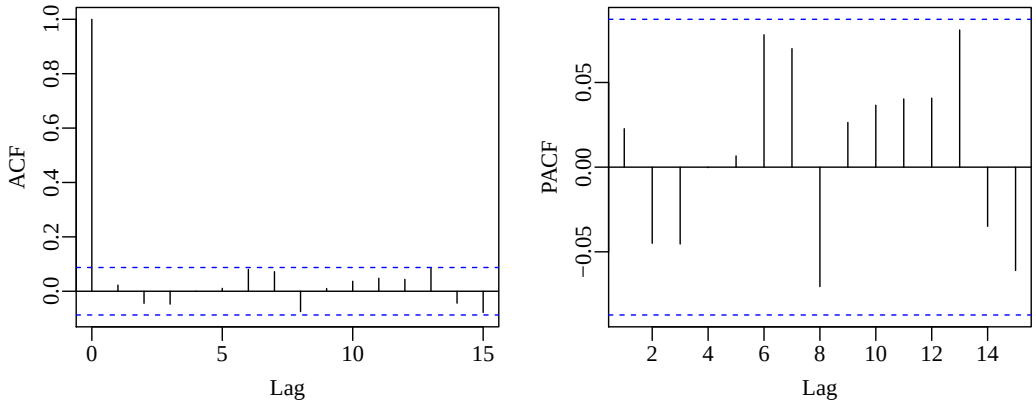
Table 2.3: *Fitted CMP-ARMA*$(1,1)$*, Negative Binomial GARMA*$(1,1)$*, and Poisson GARMA*$(1,1)$ *models for pedestrians counts data.*

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 GARMA	$\widehat{\beta}_0 = 0.5920$	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 GARMA	$\widehat{\beta}_0 = 0.5911$	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		

2.8 Conclusions

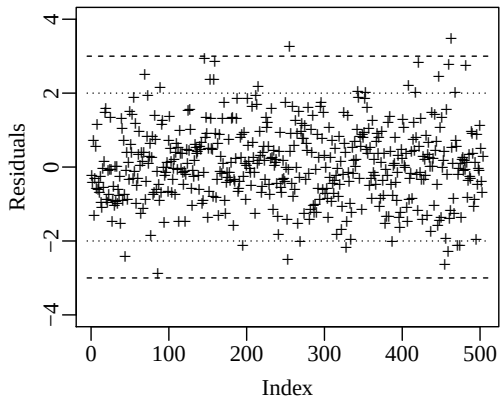
The CMP is a flexible distribution that accounts for overdispersion (or underdispersion) encountered in count data. By parametrizing the CMP distribution depending on its mean, Huang (2017) proposed a simpler and easily interpretable CMP model, while retaining all the key features of the CMP distributions that have made them increasingly attractive for the analysis of dispersed count data.

In this article, 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

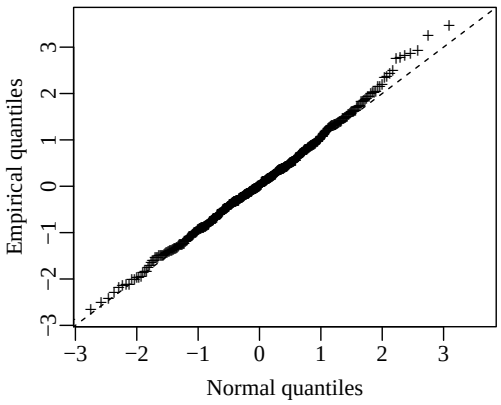


(a) Residual ACF

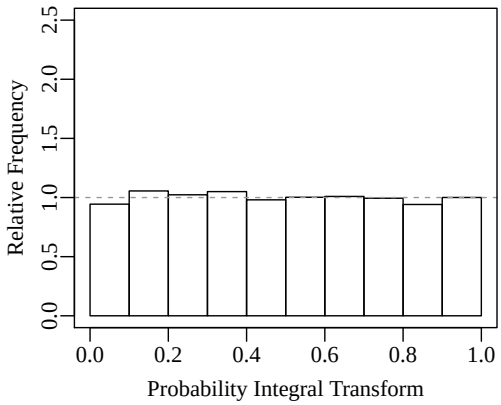
(b) Residual PACF



(c) Quantile residuals



(d) QQ-plot of the residuals

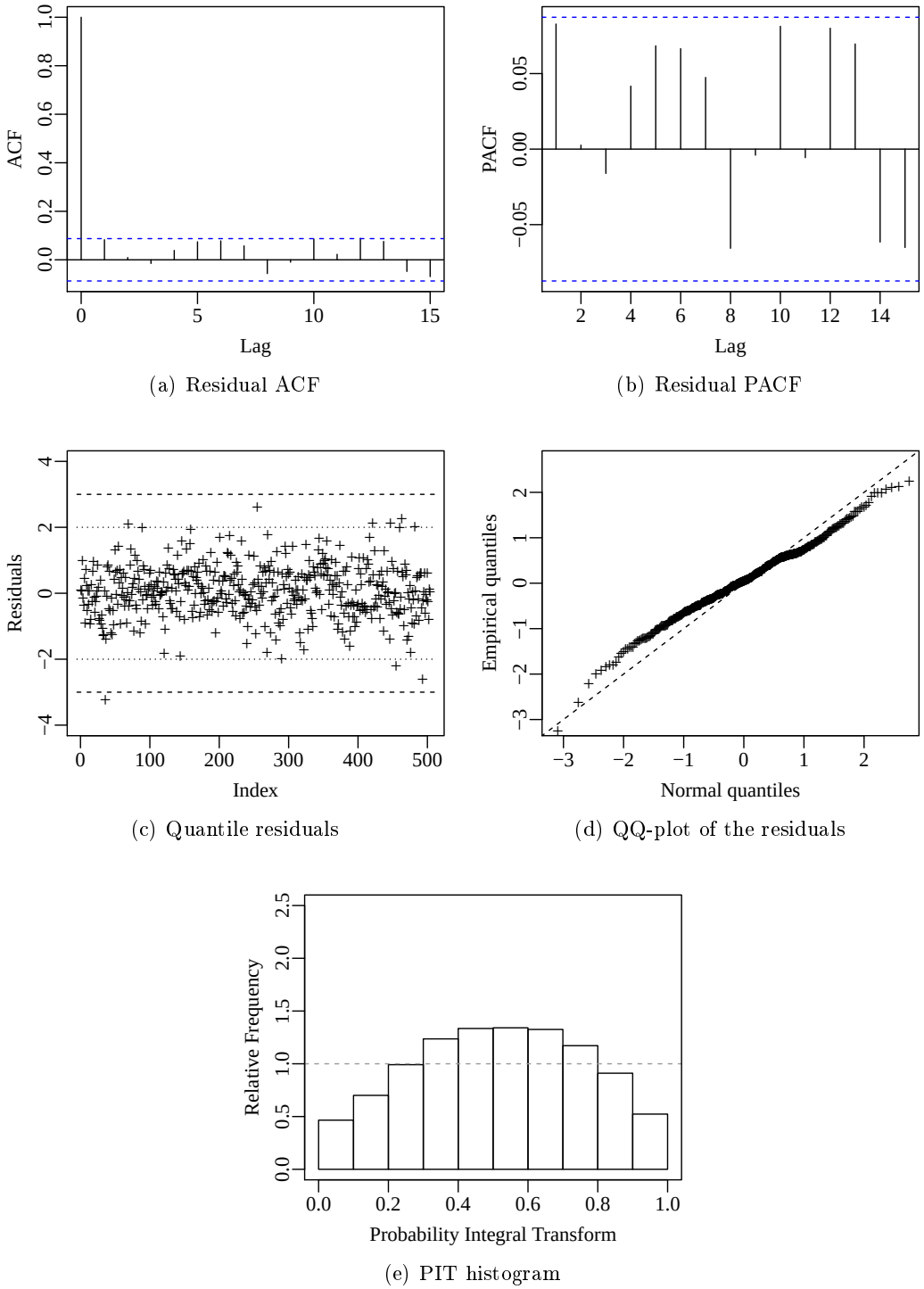


(e) PIT histogram

Figure 2.9: *Diagnostic plots for the fitted CMP-ARMA model; pedestrians counts data.*

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



(a) Residual ACF

(b) Residual PACF

(c) Quantile residuals

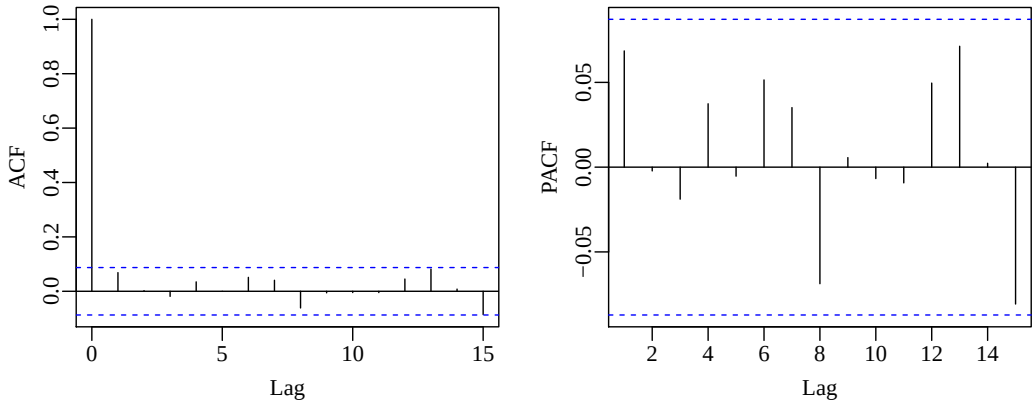
(d) QQ-plot of the residuals

(e) PIT histogram

Figure 2.10: *Diagnostic plots for the fitted Poisson GARMA model; pedestrians counts data.*

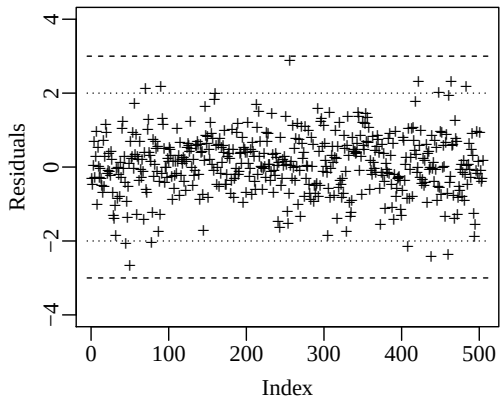
showed that the CMLE are unbiased and consistent. As in Albarracin *et al.* (2019), the GARMA model presented an overestimated moving average parameter θ while the autoregressive parameter ϕ was underestimated, indicating that there is multicollinearity between AR and MA terms. It is thus recommended 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

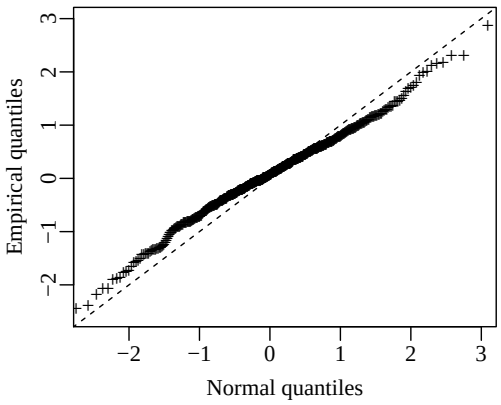


(a) Residual ACF

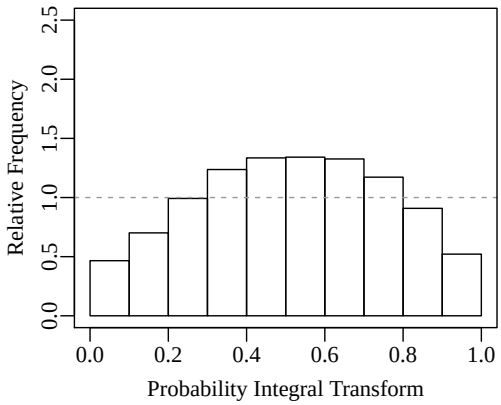
(b) Residual PACF



(c) Quantile residuals



(d) QQ-plot of the residuals



(e) PIT histogram

Figure 2.11: *Diagnostic plots for the fitted Negative Binomial GARMA model; pedestrians counts data.*

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.

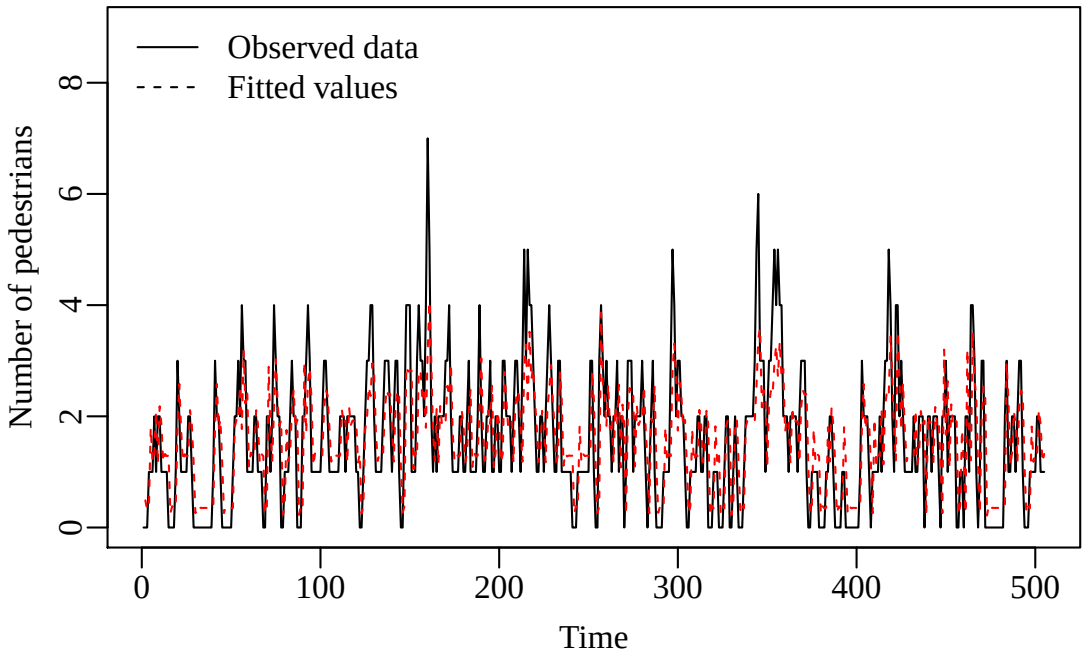


Figure 2.12: *Fitted and observed values for the CMP-ARMA(1, 1) model.*

2.9 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

Some results discussed in the Appendices A, B, and C are derived from the results presented in Huang (2017).

When $\nu = 1$, by solving (2.3) 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 (2.3) 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 (2.2) tends to ∞ for $y_t \neq 0, 1$. By solving (2.3) 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 function for $\boldsymbol{\mu_t}$

By differentiating the conditional log-likelihood function given in (2.5) 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 (2.5) 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

$$\frac{\partial \ell_t(\mu_t,\nu)}{\partial \mu_t} = \frac{\partial \ell_t(\mu_t,\nu)}{\partial \lambda(\mu_t,\nu)} \frac{\partial \lambda(\mu_t,\nu)}{\partial \mu_t} = \frac{y_t-\mu_t}{\lambda(\mu,\nu)} \frac{\lambda(\mu,\nu)}{V(\mu,\nu)} = \frac{y_t-\mu_t}{V(\mu_t,\nu)}. \tag{2.9}$$

Therefore, it follows that

$$U_{\gamma_i}(\boldsymbol{\gamma}) = \sum_{t=1}^{n} \frac{y_t-\mu_t}{V(\mu_t,\nu)g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_i}, \text{ for } \gamma_i \notin \nu. \tag{2.10}$$

Define the error in (2.4) 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 function 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 (2.5) 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

$$\frac{\partial \ell_t(\mu_t,\nu)}{\partial \nu} = \sum_{t=1}^{n} A(\mu_t,\nu)\frac{(y_t-\mu_t)}{V(\mu_t,\nu)} - \left[\log(y_t!) - B(\mu_t,\nu)\right], \tag{2.11}$$

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 (2.10) and (2.11), we then obtain the matrix expression for the score vector given in (2.6).

Appendix C - Conditional information matrix

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

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

First, $Z_t = \sum_{y_t=0}^{\infty}\lambda_t^y/(y_t!)^\nu$, 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 $V(\mu_t,\nu) = \sum\limits_{y_t=0}^{\infty}\dfrac{(y_t-\mu_t)^2\lambda^{y_t}}{(y_t!)^\nu Z_t}$, then the derivative of $V(\mu_t,\nu)$ with respect to μ_t is given by

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

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

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

Second, let $B(\mu_t,\nu) = \sum\limits_{y_t=0}^{\infty}\dfrac{\log(y_t!)\lambda_t^{y_t}}{(y_t!)^\nu Z_t}$. 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

Third, 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 (2.9), so it follows that $E\left[\frac{\partial\ell_t(\mu_t,\nu)}{\partial\mu_t}\middle|\mathcal{F}_{t-1}\right] = 0$. Therefore,

$$E\left[\frac{\partial^2\ell(\boldsymbol{\gamma})}{\partial\gamma_i\partial\gamma_j}\middle|\mathcal{F}_{t-1}\right] = \sum_{t=1}^{n}E\left[\frac{\partial^2\ell_t(\mu_t,\nu)}{\partial\mu_t^2}\middle|\mathcal{F}_{t-1}\right]\left(\frac{d\mu_t}{d\eta_t}\right)^2\frac{\partial\eta_t}{\partial\gamma_i}\frac{\partial\eta_t}{\partial\gamma_j}. \tag{2.12}$$

From (2.9) we also obtain

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

Thus,

$$E\left[\frac{\partial^2 \ell_t(\mu_t,\nu)}{\partial \mu_t^2}\middle|\mathcal{F}_{t-1}\right] = \frac{-1}{V(\mu_t,\nu)}. \tag{2.13}$$

By replacing (2.13) in (2.12) it follows that

$$E\left[\frac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}\middle|\mathcal{F}_{t-1}\right] = \sum_{t=1}^{n} \frac{-1}{V(\mu_t,\nu)g'(\mu_t)^2}\frac{\partial \eta_t}{\partial \gamma_i}\frac{\partial \eta_t}{\partial \gamma_j}. \tag{2.14}$$

From (2.14) 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 presented in the subsection 2.4.1.

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.

Chapter 3

Conway-Maxwell-Poisson seasonal autoregressive moving average model

This work proposes a new class of models, namely Conway-Maxwell-Poisson seasonal autoregressive moving average (CMP-SARMA), which extends the class of CMP autoregressive moving average models by including seasonal components to the dynamic model structure. The proposed class of models assumes a Conway-Maxwell-Poisson (CMP) conditional distribution for the variable of interest, which allows us to model univariate time series of non-negative counts with overdispersion, equidispersion, and underdispersion. As in the generalized autoregressive and moving average model (GARMA), the response's conditional mean of the proposed model is modeled by a autoregressive moving average structure. 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 conduct 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 others models in the literature.

Keywords: CMP-SARMA; Conway-Maxwell-Poisson distribution; Time series of counts; seasonality

3.1 Introduction

The generalized autoregressive moving average (GARMA) model proposed by Benjamin *et al.* (2003) has been considered in some time series applications/studies recently, see for exemple, Talamantes *et al.* (2007), Dugas *et al.* (2013), Albarracin *et al.* (2018b), and Albarracin *et al.* (2018a). 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 McCullagh and Nelder (1989), the conditional mean of the response variable is modeled directly by a regression structure through a link function.

The model proposed by Benjamin *et al.* (2003) assumes that the conditional distribution of the dependent variable belongs to the exponential family given the process history. In the same approach, Rocha and Cribari-Neto (2009) developed dynamic models in the beta distribution family

(βARMA), Bayer *et al.* (2017) introduced a dynamic class of models taking values in the double bounded interval following the Kumaraswamy distribution, and Maior and Cysneiros (2018) proposed a dynamic class of models for random variables belonging to the class of symmetric distributions. Recently, Melo and Alencar (2020) developed a dynamic regression model based on the Conway-Maxwell-Poisson (CMP) distribution for the analysis of time series of counts with equidispersion, underdispersion, and overdispersion. Although the mentioned models can be used in time series with seasonality, using sine/cosine functions as covariates, such a strategy is not appropriate when the seasonality is stochastic (Briët *et al.*, 2013).

Modeling 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 (Box *et al.*, 2015) has been frequently used for modeling univariate time series. However, many real data often do not adhere to the assumption of normality required by this model (Tiku *et al.*, 2000). Some works have shown an increasing interest in non-Gaussian seasonal time series models, such as Monteiro *et al.* (2010), Briët *et al.* (2013), Bourguignon *et al.* (2016), and Bayer *et al.* (2018). However, the study of seasonal time series of counts with equidispersion, underdispersion, and overdispersion has received less attention in the literature. Thus, this article 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 Huang (2017) that allows the mean to be modeled directly. The model parameters are estimated using the conditional maximum likelihood estimation. In addition, some residual and diagnostic tools are proposed and discussed. Finally, a Monte Carlo simulation study is carried to evaluate the estimators performance.

This article is organized as follows. Initially, we present the proposed model in Section 3.2. The parameters of the model are estimated by the conditional maximum likelihood method in Section 3.3, where we also provide closed-form expressions for the conditional score vector and conditional Fisher information matrix. The discussion in Section 3.3 also provides confidence intervals and hypothesis testing. Section 3.4 gives some diagnostic measures and forecasting. In Section 3.5, we conduct a brief simulation study. Section 3.6 presents an empirical application to illustrate the proposed model. Finally, Section 3.7 gives some conclusions.

3.2 The proposed model

Let $Y = (Y_1, \ldots, Y_n)^\top$ denote the n-dimensional vector of random variables and let $\mathcal{F}_{t-1} = \{Y_{t-1}, \ldots, Y_1\}$ denote previous information set. In the CPM-SARMA model, the conditional distribution of Y_t, given $\mathcal{F}_{t-1}$, is assumed to follow the CMP distribution with mean μ_t and dispersion ν, that is, the conditional probability is given by (Huang, 2017):

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

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

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

$Z(\lambda(\mu_t, \nu), \nu) = \sum_{s=0}^{\infty} \lambda(\mu_t, \nu)^s / (s!)^\nu$ is a normalization function, and 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 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

$$\phi(B)\Phi(B^S)[g(y_t) - x_t^\top \boldsymbol{\beta}] = \theta(B)\Theta(B^S) r_t, \tag{3.1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_r)^\top$ is the $(r+1)$-dimensional vector of unknown parameters, $\boldsymbol{x}_t = (1, x_{1,t}, \ldots, x_{p,t})^\top$ is the $(r+1)$-dimensional vector containing the covariates at time t, $r_t = g(y_t) - g(\mu_t)$ is the error term, $g(\cdot)$ 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 the polynomials

$$\begin{aligned}
\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \\
\theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q, \\
\Phi(B^S) &= 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \cdots - \Phi_P B^{SP}, \\
\Theta(B^S) &= 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \cdots - \Theta_Q B^{SQ},
\end{aligned}$$

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

We can rewrite Equation (3.1) as

$$\begin{aligned}
\eta_t = g(\mu_t) = x_t^\top \boldsymbol{\beta} &+ \sum_{i=1}^{p} \phi_i [g(y_{t-i}) - x_{t-i}^\top \boldsymbol{\beta}] + \sum_{I=1}^{P} \Phi_I [g(y_{t-IS}) - x_{t-IS}^\top \boldsymbol{\beta}] \\
&- \sum_{i=1}^{p} \sum_{I=1}^{P} \phi_i \Phi_I [g(y_{t-(i+IS)}) - x_{t-(i+IS)}^\top \boldsymbol{\beta}] - \sum_{j=1}^{q} \theta_j r_{t-j} - \sum_{J=1}^{Q} \Theta_J r_{t-(JS)} \\
&+ \sum_{j=1}^{q} \sum_{J=1}^{Q} \theta_j \Theta_J r_{t-(j+JS)},
\end{aligned} \tag{3.2}$$

where $\eta_t = g(\mu_t)$ is the linear predictor. 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, and to avoid calculating the logarithm of observations equal to zero, we replace y_{t-j} in Equation (3.1) for $y_{t-j}^* = \max(y_{t-j}, c)$, thus replacing $y_{t-j} = 0$ by an arbitrary small value c such that $0 < c < 1$. The structure of the model is the same as in Bayer *et al.* (2018)

The CMP-SARMA model extends the approach proposed by Melo and Alencar (2020) by incorporating a seasonal autoregressive moving average (SARMA) structure.

3.3 Parameter estimation

The parameter estimation of the CMP-SARMA model can be obtained by 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

$$\ell(\boldsymbol{\gamma}) = \sum_{t=m+1}^{n} \log Pr(y_t|\mathcal{F}_{t-1}) = \sum_{t=m+1}^{n} \ell_t(\mu_t, \nu), \tag{3.3}$$

where

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

3.3.1 Conditional score vector

The conditional score vector is given by taking first derivatives of the conditional log-likelihood function given in (3.3) with respect to each element of $\boldsymbol{\gamma}$. In what follows, we have

$$U_{\gamma_i}(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_i} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \nu)}{\partial \lambda(\mu_t, \nu)} \frac{\partial \lambda(\mu_t, \nu)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_i}. \tag{3.4}$$

Observe that

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

where $V(\mu_t, \nu) = \sum\limits_{y_t=0}^{\infty} \dfrac{(y_t - \mu_t)^2 \lambda_t(\mu_t, \nu)_t^y}{(y_t!)^\nu Z(\lambda_t(\mu_t, \nu), \nu)}$ is the variance of Y_t. Substituting these results in (3.4), we obtain

$$U_{\gamma_i}(\boldsymbol{\gamma}) = \sum_{t=1}^{n} \frac{y_t - \mu_t}{V(\mu_t, \nu) g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_i}, \text{ for } \gamma_i \notin \nu. \tag{3.5}$$

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 $\boldsymbol{Z}$, $\boldsymbol{A}$, $\boldsymbol{\mathcal{A}}$, $\boldsymbol{M}$, and $\boldsymbol{\mathcal{M}}$ 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

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

For each $\gamma_i \notin \nu$ in Equation (3.5), each element can be written in matrix form as

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

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 $\boldsymbol{U}(\boldsymbol{\gamma}) = \boldsymbol{0}$, where $\boldsymbol{0}$ is a vector of zeros with dimension $(r+p+P+q+Q+2)$. Such a system does not have an analytical solution, being necessary to apply iterative numerical methods. Here, we apply the `nlminb` optimization algorithm (Gay, 1990) available from the `stats` package in `R` software (R Core Team, 2019). 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 (3.2).

3.3.2 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 condi-

tional 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 $i, j \in \{1, \ldots, r+p+P+q+Q\}$ $(\gamma_i \neq \nu,)$, is given by

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

Since $E\left(\frac{\partial\ell_t(\mu_t,\nu)}{\partial\mu_t}\middle|\mathcal{F}_{t-1}\right) = 0$ (Melo and Alencar, 2020) the expected value of the derivative above is given by

$$E\left(\frac{\partial^2\ell(\boldsymbol{\gamma})}{\partial\gamma_i\partial\gamma_i}\middle|\mathcal{F}_{t-1}\right) = \sum_{t=m+1}^{n} E\left(\frac{\partial^2\ell_t(\mu_t,\nu)}{\partial\mu_t^2}\middle|\mathcal{F}_{t-1}\right)\left(\frac{d\mu_t}{d\eta_t}\right)^2\frac{\partial\eta_t}{\partial\gamma_i}\frac{\partial\eta_t}{\partial\gamma_j}. \tag{3.6}$$

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

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

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

$$E\left(\frac{\partial^2\ell_t(\mu_t,\nu)}{\partial\mu_t^2}\middle|\mathcal{F}_{t-1}\right) = \frac{-1}{V(\mu_t,\nu)}. \tag{3.7}$$

By replacing (3.7) in (3.6), it follows that

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

The second order derivatives of the log-likelihood function related to ν are given by (Melo and Alencar, 2020)

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

and

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

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

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

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

where

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

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 (Andersen, 1970; Bayer *et al.*, 2017; Pumi *et al.*, 2019), that is,

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

where $N_{r+p+q+P+Q+2}$ denotes the $(r+p+q+P+Q+2)$-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 $\boldsymbol{K}(\boldsymbol{\gamma})^{-1}$ is the conditional Fisher information inverse matrix.

3.3.3 Confidence intervals and hypothesis testing

Considering 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 (Pawitan, 2001) is the so-called z statistic, which is given by

$$z = \frac{\widehat{\gamma}_i - \gamma_i}{\sqrt{k^{ii}}},$$

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 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 Kedem and Fokianos (2005) 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 density function of the standardized normal distribution $N(0,1)$.

3.4 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 use a modification of the Akaike Information Criterion (AIC) (Akaike, 1974) for dynamic models proposed by Bayer *et al.* (2018), which is given by

$$\text{MAIC} = -\widehat{\ell}_* + 2k, \tag{3.8}$$

where $\widehat{\ell}_* = \frac{n}{n-m}\ell(\widehat{\gamma})$ and $k = r+p+q+P+Q+1$ is the number of parameters in the model. The models can be selected using the modified Schwarz Information Criterion (MSIC) (Schwarz *et al.*, 1978), obtained by replacing the term $2k$ by $\log(n)k$ in (3.8).

As an alternative, we can also to evaluate the goodness-of-fit of the models by the deviance, given by $D = 2(\ell(\tilde{\gamma}) - \ell(\widehat{\gamma}))$, where $\ell(\tilde{\gamma})$ and $\ell(\widehat{\gamma})$ are the conditional log-likelihood function of the saturated ($\widehat{\mu}_t = y_t$) and fitted models, respectively. When the fitted model is well adjusted, the test statistic D has the asymptotic chi-squared distribution with $n-(r+p+q+P+Q+1)$ degrees of freedom (Benjamin *et al.*, 2003; Fokianos and Kedem, 2004). If $D/n-(r+p+q+P+Q+1) \gg 1$, then the fit to the data can be considered inadequate.

Residual analysis is an important technique for checking model adequacy. Although there are several ways for specifying residuals, the traditional residuals are typically not normally distributed given the true model (Feng *et al.*, 2017).

Here, we consider the randomized quantile residuals introduced by (Dunn and Smyth, 1996). When a response variable is discrete the randomized quantile residuals is 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 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 distribution.

To test the validity of the assumed distribution in the proposed model, we use a non-randomized yet uniform version of the probability integral transformation (PIT) proposed by Czado *et al.* (2009) for time series models for count data. The non-randomized PIT is given as follows

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

where $F^{(t)}(u \mid \mathcal{F}_{t-1})$ is the conditional cumulative distribution function of observed counts y_t, given by

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

The PIT has a standard uniform distribution when the model is correctly specified (Jung and Tremayne, 2011; Jung *et al.*, 2016).

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

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

where

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

3.5 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

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

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 article were implemented and performed in R (R Core Team, 2019). We generate 5,000 Monte Carlo replicates of each experiment with $n \in \{100, 200, 400, 800\}$. For each experiment, we evaluate the mean, percentage relative bias (RB %), defined as $\{E(\widehat{\gamma}_i) - \gamma_i\}/\gamma_i$, and mean squared error (MSE).

The results for all scenarios are shown in Table 3.1. We note that the bias decreases and that the MSE tend toward zero as the size of the sample increases, indicating the consistency property of the CMLE. This results indicate that the CMLE appeared to perform well. We also note that the seasonal estimators present a larger relative bias in all scenarios. Such a fact was also verified for the βSARMA model in Bayer *et al.* (2018).

3.6 Empirical application

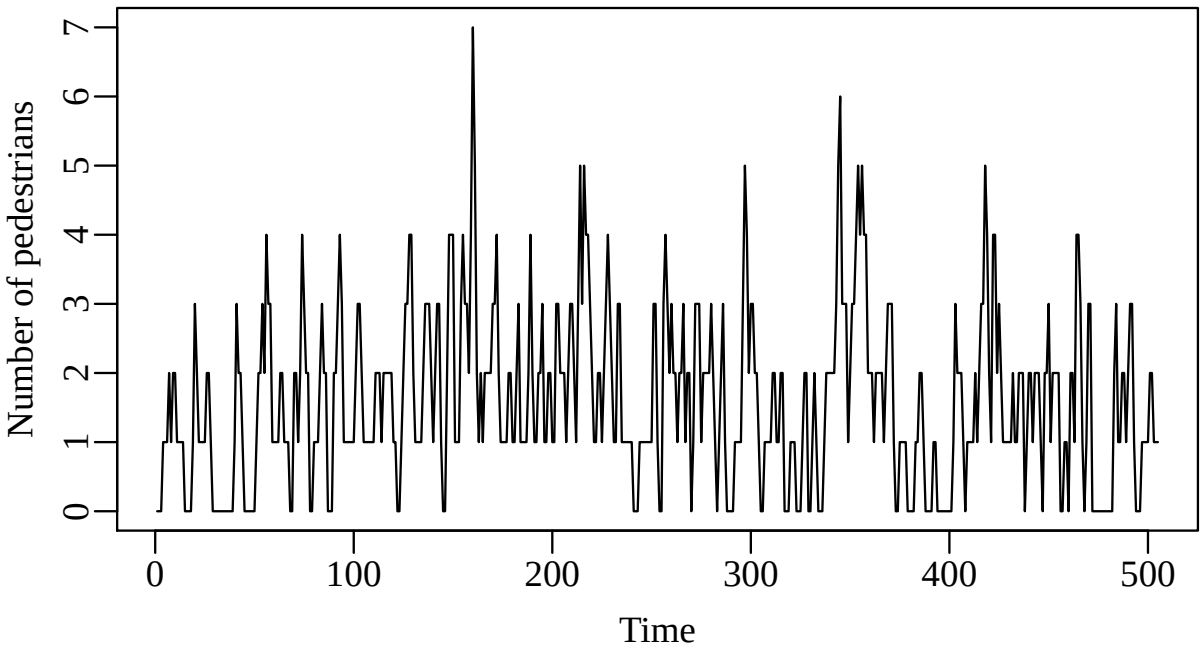
In this section, for illustrative purposes, we analyze and discuss a real data application. The data consists of 120 observations of the monthly number of claims of short-term disability benefits made by the Workers' Compensation Board of British Columbia, from January 1985 to December 1994. All the claimants are male, between the ages of 35 and 54 years, work in the logging industry, and have injuries caused by cuts, lacerations, and punctures.

This series was originally studied by Freeland (1998), who considered an INAR(1) model using sine/cosine functions as covariates to explain the seasonality of this data set. Recently, Bourguignon *et al.* (2016) fitted a Poisson seasonal INAR$(1)_s$ model to this same data set. Figure 3.1 presents the time series plot and its sample autocorrelation function (ACF) and partial autocorrelation function (PACF). These plots indicate the presence of a seasonal dynamic.

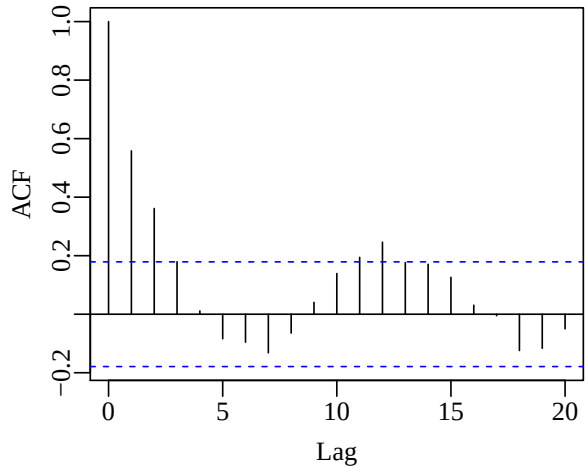
For comparison purposes, we also fitted the CMP-ARMA and negative binomial GARMA (NB-GARMA) models to these data. For these models, the seasonal effect is modeled including deterministic cosine and sine functions with an annual cycle, $x_t = (\cos(2\pi t/12), \sin(2\pi t/12))$, $t = 1, \ldots, n$. Here, we consider the parameterization given in Evans (1953), where 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; and σ is called the dispersion parameter. The NB-GARMA model was fitted using the

Table 3.1: *Monte Carlo simulation results of the CMLE for the CMP-ARMA* $(1,1) \times (1,1)_{12}$ *model.*

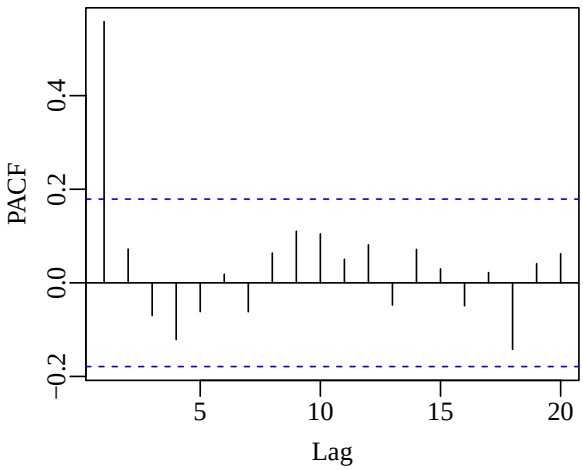
Scenario 1 - overdispersion							
		β_0	ϕ_1	Φ_1	θ_1	Θ_1	ν
	Parameters	2.0	0.5	−0.2	−0.4	0.3	0.5
	Mean	2.299	0.469	−0.300	−0.413	0.211	0.539
$n = 100$	RB(%)	14.966	−6.113	50.058	3.366	−29.501	7.937
	MSE	0.506	0.018	0.040	0.019	0.055	0.018
	Mean	2.161	0.486	−0.261	−0.406	0.249	0.518
$n = 200$	RB(%)	8.043	−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							
		α	ϕ_1	Φ_1	θ_1	Θ_1	ν
	Parameters	2.0	0.5	−0.2	−0.4	0.3	1.0
	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							
		α	ϕ_1	Φ_1	θ_1	Θ_1	ν
	Parameters	2.0	0.5	−0.2	−0.4	0.3	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



(a) Time series



(b) Sample ACF



(c) Sample PACF

Figure 3.1: *Time series, ACF and PACF plots for the monthly number of claims time series from* 1985 *to* 1994.

garmaFit function from *gamlss.util* (Stasinopoulos and Rigby, 2016) library in the R software.

Based on the previously mentioned diagnostic measures, the CMP-SARMA$(1,0) \times (1,0)_{12}$, CMP-ARMA$(1,0)$, and NB-GARMA$(1,0)$ models were selected. The parameter estimates with corresponding standard errors (shown in parentheses), and MAIC and MSIC values for the selected models are given in Table 3.2. Notice that the CMP-SARMA model presented the smallest MAIC and MSIC values, suggesting that the proposed model yielded a better fit to the data than the other two non-seasonal models. The estimated dispersion parameter indicates overdispersion ($\nu < 1$).

Figure 3.2 presents a diagnostic analysis of the fitted CMP-SARMA$(1,0) \times (1,0)_{12}$ model. The ACF and PACF of randomized quantile residuals in Figures 3.2(a) and 3.2(b), respectivel,

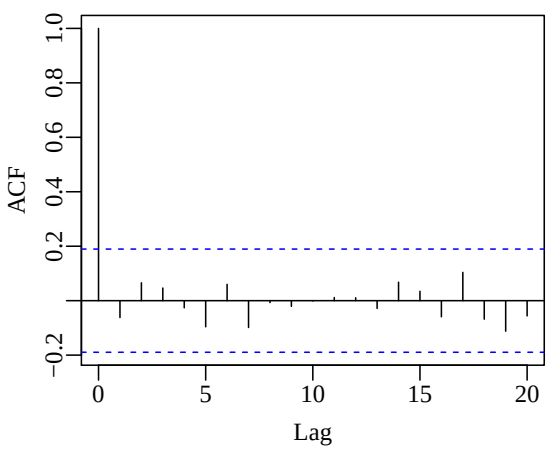
Table 3.2: *Parameter estimates, standard errors (shown in parentheses), and model selection criteria; monthly count of claims data*

Model		Estimate(se)	MAIC	MSIC
CMP-SARMA$(1,0)\times(1,0)_{12}$	$\widehat{\beta}_0$	1.9353(0.1128)	567.711	578.861
	$\widehat{\phi}_1$	0.5149(0.0761)		
	$\widehat{\Phi}_1$	0.1905(0.0886)		
	$\widehat{\nu}$	0.7932(0.1222)		
CMP-ARMA$(1,0)$	$\widehat{\beta}_0$	1.8398(0.0742)	569.455	583.376
	$\widehat{\beta}_1$	−0.1871(0.0853)		
	$\widehat{\beta}_2$	−0.2819(0.0877)		
	$\widehat{\phi}_1$	0.4526(0.0761)		
	$\widehat{\nu}$	0.8860(0.1270)		
NB-GARMA$(1,0)$	$\widehat{\beta}_0$	1.8384(0.0754)	568.833	582.771
	$\widehat{\beta}_1$	−0.1889(0.0870)		
	$\widehat{\beta}_1$	−0.2839(0.0896)		
	$\widehat{\phi}_1$	0.4489(0.0774)		
	$\widehat{\alpha}$	0.0263(0.0232)		

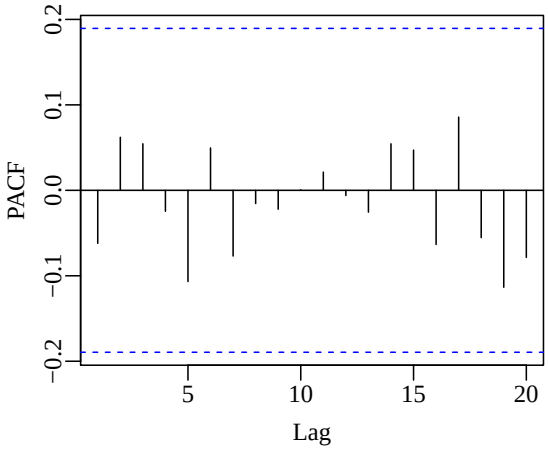
indicate that there is no significant autocorrelation in the residuals, which is confirmed by the Ljung-Box Q test (Ljung and Box, 1978) based on 15 lags. The ACF and PACF of randomized quantile residuals are presented in Figures 3.2(a) and 3.2(b), respectively. Notice that there is no indication of significant autocorrelation in the residuals, which is confirmed by the Ljung-Box Q test (Ljung and Box, 1978) based on 15 lags. The Ljung-Box test does not reject the null hypothesis with p-value $= 0.9794$. By looking at the residual plot in Figure 3.2(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 3.2(d). The uniformity of the PIT in Figure 3.2(e) (with $J = 10$) indicates that the model was correctly adjusted. Finally, Figure 3.3 shows the observed and fitted values for the CMP-ARMA$(1,0)\times(1,0)_{12}$ model.

3.7 Conclusions

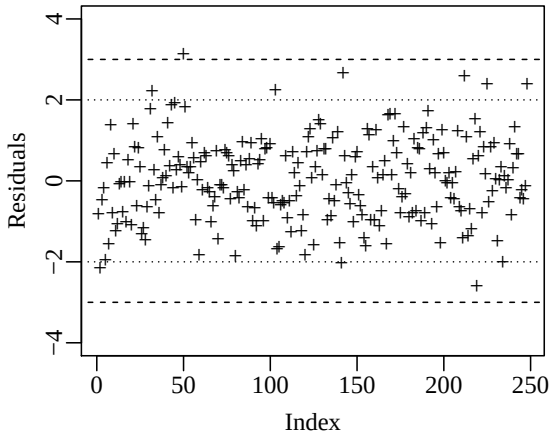
The present work proposed the class of Conway-Maxwell-Poisson seasonal autoregressive moving average model CMP-SARMA $(p,q)\times(P,Q)_s$ models for time series of counts. This class of models includes seasonal components in which the class of CMP-ARMA models is a special case, and the CMP-SARMA model can be used to model overdispersed, equidispersion, and underdispersed data. We assumed that the conditional distribution of the response variable follows a CMP distribution (Huang, 2017). 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 Gaussian for finite samples. 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 error is white noise. Finally, we pre-



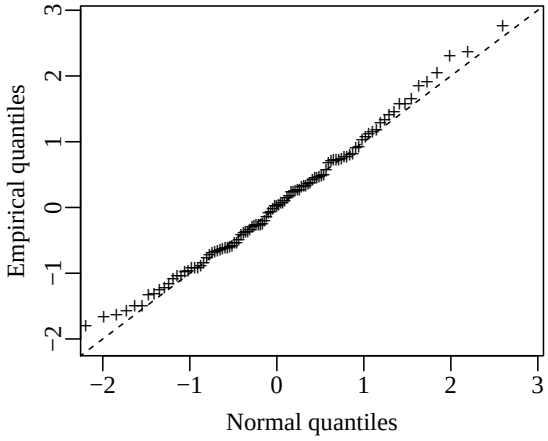
(a) Residual ACF



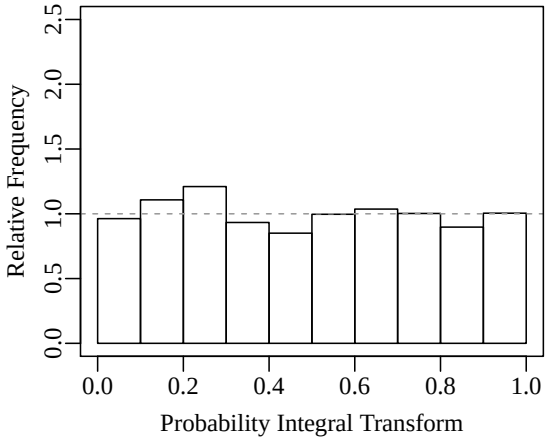
(b) Residual PACF



(c) Quantile residuals



(d) QQ-plot of the residuals



(e) PIT histogram

Figure 3.2: *Diagnostic plots for the fitted CMP-SARMA model; monthly counts of claims data.*

sented and investigated an empirical application that illustrated the usefulness and applicability of the proposed model.

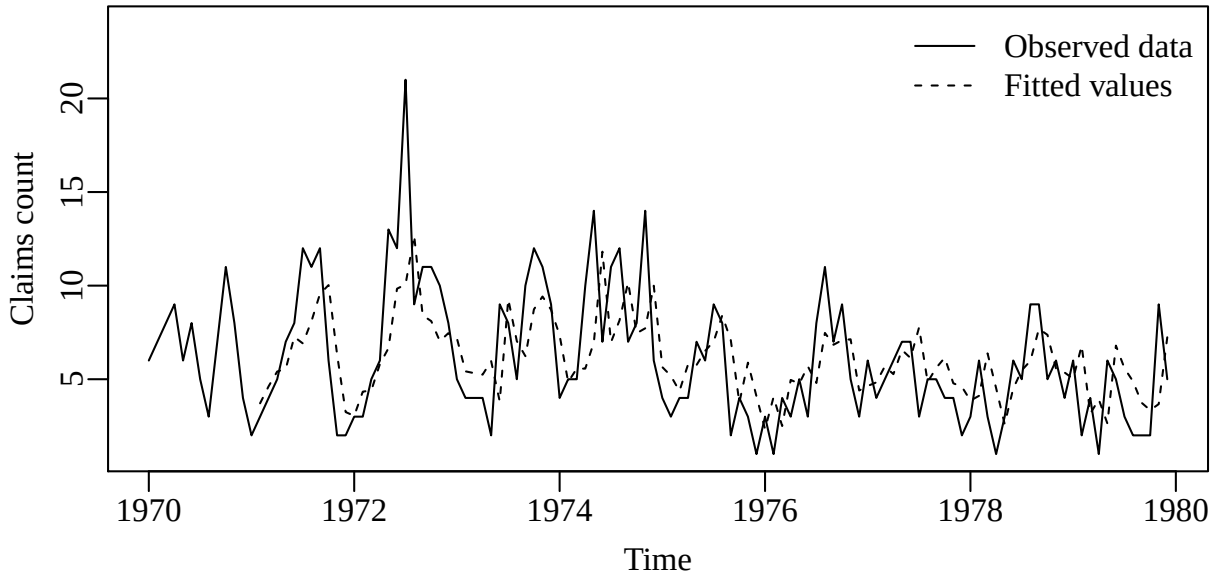


Figure 3.3: *Fitted and observed values; monthly count of claims data.*

Chapter 4

Progressive Mean Control Chart for Monitoring Count Time Series

Several control charts have been proposed to assist the monitoring of public health disease outbreaks. Memory-type control charts are commonly used to monitor these data because they efficiently detect small shifts. In this context, we propose a new memory-type control chart, in which a progressive mean is used as the plotting statistic for monitoring autocorrelated count data, typically arising in health surveillance. The development of the new control chart is based on the randomized quantile residuals obtained from a fitted Conway-Maxwell-Poisson autoregressive moving average (CMP-ARMA) model. We carry out a Monte Carlo simulation study to evaluate and compare the performance of the proposed control chart with two traditional approaches in quality control; Shewhart-type and exponentially weighted moving average (EWMA) control charts. The results show that the proposal outperforms the two conventional control charts considered. Finally, we illustrate the applicability of the proposed control chart by monitoring the weekly number of hospitalizations of people older than 60 years due to respiratory diseases in the city of São Paulo, Brazil.

Keywords: Control charts; Progressive mean; Randomized quantile residual; Average run length; CMP-ARMA

4.1 Introduction

Control charts are often used to signal the occurrence of special causes that can affect the quality of industrial and service processes (Montgomery, 2007). Shewhart (1927, 1926) first introduced control charts to monitor production processes, and since then, they have been applied in several fields, such as medicine, marketing, analytical laboratories, health care, and engineering. In public health surveillance, we are usually interested in detecting changes in the number of cases of a specific disease in order to decide if this morbidity or mortality reached an epidemic level (Albarracin *et al.*, 2018b).

Shewhart-type control charts are characterized by using only the current process information, ignoring the previous behavior of the process. For this reason, they are known as memory-less control charts. The typical Shewhart chart is known to be sensitive only for large process shifts (Saghir and Lin, 2014c). On the other hand, memory-type control charts, such as the cumulative

sum (CUSUM) (Page, 1954) and exponentially weighted moving average (EWMA) (Roberts, 1959) control charts, use both the past and current information, which makes them more efficient in detecting small to moderate shifts in the process (Abbasi and Miller, 2013).

Modifications of these two memory-less charts have been proposed in the literature. See, for instance, Capizzi and Masarotto (2003), Jiang *et al.* (2008), Shu and Jiang (2008), Abbas *et al.* (2011), Abbasi *et al.* (2012), Haq *et al.* (2014), and Haq *et al.* (2015). Recently, Abbas *et al.* (2013) proposed a new memory-type control chart for monitoring the mean of the process using individual observations. This chart is based on the progressive mean (PM) statistic, which can be viewed as an adaptive EWMA statistic where the smoothing parameter is updated after every sample (Abbas, 2015). Based on the average run length (ARL) measure, defined as the average number of samples until a signal, Abbas *et al.* (2013) showed that the PM control chart better detects small and moderate shifts in the process, compared to the Shewhart, EWMA, and CUSUM charts. In addition, the PM control chart shows good performance for large shifts. Recently, several control charts based on the PM statistic have been proposed in the literature. See, for instance, Abbasi *et al.* (2013), Abbasi (2017), Abbasi *et al.* (2019), and Alevizakos and Koukouvinos (2019b).

As already mentioned, the progressive mean statistic has received considerable attention in recent years. Despite this, to the best of our knowledge, applications of this statistic in control charts for monitoring count time series models have never been considered in the literature. Therefore, in the present article, we propose a new memory-type control chart based on the progressive mean statistic for monitoring the count series that may present underdispersion, equidispersion, or overdispersion. The new chart is used to monitor the time-varying mean response using randomized quantile residuals from a fitted Conway-Maxwell-Poisson autoregressive moving average (CMP-ARMA) model. Several control charts have been developed based on the CMP distribution. See, for example, Sellers (2012), Saghir and Lin (2014a), Saghir and Lin (2014b), Aslam *et al.* (2016), Aslam *et al.* (2017), Aslam *et al.* (2018), Alevizakos and Koukouvinos (2019a), Alevizakos and Koukouvinos (2019b), and Rao *et al.* (2020).

Although several types of residuals can be employed to evaluate these classes of models, we consider the randomized quantile residual (Dunn and Smyth, 1996) because the Pearson and deviance residuals may be far from normality in count data with low fitted means (Benjamin *et al.*, 2003). Recently, Park *et al.* (2020) proposed Shewhart-type control charts based on randomized quantile residuals for a wide range of the response variable and Albarracin *et al.* (2018a) introduced a like-EWMA control chart to monitor autocorrelated count series based on conditional quantile residuals from a generalized autoregressive moving average (GARMA) model.

In this article, we evaluate and compare the performance of the proposed control chart with the Shewhart-type and EWMA control charts through simulation studies. The performance is measured in terms of ARLs. The rest of the article is organized as follows. Section 4.2 presents a brief review of CMP-ARMA models. Section 4.3 introduces randomized quantile residuals and details the construction of the new control chart. In Section 4.4, we conduct a Monte Carlo simulation study to evaluate the performance of the proposal in terms of run length characteristics. Section 4.5 presents a comparison study of the proposed chart with the Shewhart-type control chart proposed by Park *et al.* (2020) and the traditional EWMA chart (Roberts, 1959) in terms of ARL. In Section 4.6, the new control chart is used for monitoring the weekly number of hospital admissions due to respiratory diseases of people older than 60 years in São Paulo city. Conclusions and future

directions are discussed in Section 4.7.

4.2 Conway-Maxwell-Poisson autoregressive moving average model

In this section, we shall present the CMP-ARMA model proposed by Melo and Alencar (2020) for modeling underdispersed, equidispersed, and overdispersed count time series data. This class of models is very flexible and contains the GARMA models with the Poisson, Geometric, and Bernoulli distributions as particular cases.

Let $Y = (Y_1, \ldots, Y_n)^\top$ be a vector of n random variables and assume that the conditional distribution of each observation y_t, conditioned on the past information set $\mathcal{F}_{t-1} = \{Y_{t-1}, \ldots, Y_1\}$, follows a CMP distribution with conditional probability function given by

$$Pr(Y_t = y_t \mid \mathcal{F}_{t-1}, \mu_t, \nu) = \frac{\lambda(\mu_t, \nu)^{y_t}}{(y_t!)^\nu Z(\lambda(\mu_t, \nu), \nu)}, \quad y_t = 0, 1, 2, \ldots, \tag{4.1}$$

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

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

and $Z(\lambda(\mu_t, \nu), \nu) = \sum_{s=0}^{\infty} \lambda(\mu_t, \nu)^s / (s!)^\nu$ is a normalization function, and the conditional mean of Y_t is given by $E(Y_t \mid \mathcal{F}_{t-1}) = \mu_t$. The parameterization in (4.1) was proposed by Huang (2017), which allows us to model the mean directly, making it possible for the construction of simpler and more interpretable models.

The CMP-ARMA model is obtained by assuming a linear predictor for the conditional mean of y_t with the following structure

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

where α is an intercept, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)^\top$ 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.

Note that in the CMP-ARMA model, as in the GARMA models proposed by Benjamin *et al.* (2003), the conditional mean μ_t is related to the linear predictor containing autoregressive and moving average terms besides time-varying regressors.

4.3 The PM Control Chart for Regression Models

This section is divided into two subsections. Subsection 4.3.1 presents a brief review of the randomized quantile residual and Subsection 4.3.2 presents the construction of the proposed control chart.

4.3.1 Randomized quantile residual

As already mentioned, the distribution of the deviance and Pearson residuals are usually non-normally distributed for discrete data with low fitted means. For this reason, Melo and Alencar (2020) recommend the randomized quantile residual for CMP-ARMA models. Feng *et al.* (2017) showed theoretically that the randomized quantile residual follows a standard normal distribution under the true model. Despite this, this residual has received little attention in the literature.

Let $a_t = F(y_t - 1 \mid \widehat{\theta})$ and $b_t = F(y_t \mid \widehat{\theta})$, where F is the fitted cumulative distribution function. For the discrete distribution function, the randomized quantile residual for y_t is defined as

$$r_t^{(q)} = \Phi^{-1}(u_t), \tag{4.2}$$

where Φ^{-1} is the quantile function of the standard normal distribution and u_t 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.

4.3.2 Designing of the proposed control chart

In this section, we shall present the construction of the proposed control chart. Let $r_1^q, \ldots, r_t^q$ be the randomized quantile residuals of the fitted model. The plotting statistic for the PM chart is given by

$$PM_t = \frac{\sum_{i=1}^{t} r_i^{(q)}}{t}. \tag{4.3}$$

As the randomized quantile residuals follow a standard normal distribution under the true model, we have that the mean and variance of the PM_t statistic are given, respectively, by

$$\begin{aligned} \mathbb{E}(PM_t) &= \mu_0, \\ \mathrm{Var}\,(PM_t) &= \frac{\sigma_0^2}{t}. \end{aligned}$$

Therefore, according to the typical Shewhart control limits, the upper control limit (UCL) and the lower control limit (LCL) are defined as follows

$$\begin{aligned} UCL_t &= \mu_0 + L\sqrt{\frac{\sigma_0^2}{t}}, \\ CL_t &= \mu_0, \\ LCL_t &= \mu_0 - L\sqrt{\frac{\sigma_0^2}{t}}, \end{aligned} \tag{4.4}$$

where L is a constant that defines the width of the control limits. Note that the control limits given in Equation (4.4) vary over time. Although the control limits become narrower as t increases, they are too wide relative to the plotting statistic for large values of t, which results in a tiny probability of detecting a point beyond the control limits. In order to remedy this problem, Abbas *et al.* (2013) included a function that penalizes the control limits, that is, larger values of t are associated with

more penalized limits. The penalized limits for the proposed PM_t chart are given by

$$UCL_t = \mu_0 + L\frac{1}{f(t)}\sqrt{\frac{\sigma_0^2}{t}},$$
$$CL_t = \mu_0,$$
$$LCL_t = \mu_0 - L\frac{1}{f(t)}\sqrt{\frac{\sigma_0^2}{t}},$$

where $f(t)$ is an arbitrary function of t used to control the run lengths (Abbas *et al.*, 2013). We considered $f(t) = t^{0.2}$, as in Abbas *et al.* (2013), Abbasi *et al.* (2013), Abbasi (2017), and Alevizakos and Koukouvinos (2019b).

4.4 Performance evaluation

In this section, we investigate the run length distribution and evaluate the performance of the proposed control chart in terms of ARL. To compare the ARL of the out-of-control process, ARL_1, we consider that the mean parameter is shifted to $\mu_{1,t} = \mu_{0,t}(1 \pm d)$, with $d \in \{0.05, 0.1, 0.2, 0.4, 0.8\}$, where $\mu_{0,t}$ and $\mu_{1,t}$ are the mean of the y_t when the process is in control and out of control, respectively. We ran 5000 Monte Carlo replications. Time series are simulated using a CMP-ARMA model with mean

$$\eta_t = \log(\mu_t) = \phi_1 \log(y_{t-1}) + \phi_2 \log(y_{t-2}) + \theta_1[\log(y_{t-1}) - \eta_{t-1}], \tag{4.5}$$

where $\phi_1 = 0.5$, $\phi_2 = -0.3$, and $\theta_1 = 0.5$. We considered three different values of $\nu \in \{0.5, 1.0, 2.0\}$, that is, overdispersion, equidispersion, and underdispersion, respectively.

4.4.1 The run-length distribution

To evaluate the run length distribution and the performance of the proposed control chart, we conducted an extensive Monte Carlo simulation study using the `R` software (R Core Team, 2019). This simulation study is described in the following steps.

1. Phase I

   a. Select an initial value for L.

   b. Generate time series of size $n = 3000$ from the CMP-ARMA model with in-control mean, $\mu_{0,t}$, given by Equation (4.5).

   c. Fit the CMP-ARMA model using the simulated observations and obtain the randomized quantile residuals using Equation (4.2).

   d. Compute the PM_t statistic as given in Equation (4.3).

   e. Calculate the number of samples until we get the first out-of-control signal (run length). The process is declared to be in control if $LCL_t \leq PM_t \leq UCL_t$. Otherwise, declare the process as out-of-control.

f. Repeat the above steps 5000 times to obtain the mean of the run length (ARL). Compare the ARL value obtained with the fixed ARL_0 value. If the value found is equal or approximately equal to the specified ARL_0 value, then go to Phase II. Otherwise, go back to Step *a*.

2. Phase II

a. Generate time series of size $n = 3000$ from the CMP-ARMA model with parameter $\mu_{1,t} = \mu_{0,t}(1 \pm d)$, where d is the mean shift to be detected.

b. Fit the CMP-ARMA model using the simulated observations and obtain the randomized quantile residuals using Equation (4.2).

c. Compute the PM_t statistic as given in Equation (4.3).

d. Calculate the number of samples until we get the first out-of-control signal.

e. Repeat the above steps 5000 times to obtain the mean ARL_1, standard deviation (SDRL), and percentile points of the run length distribution.

Table 4.1 presents a summary of the run length characteristics for the proposed control chart. We observe that the new control chart efficiently detects small, moderate, and large shifts in the process mean. In addition, when the shift increases, the ARL, SDRL, and percentile points of the run length distribution decreases, as expected. We also notice that the run length distribution of the proposed chart is positively skewed. Note that the proposed control chart detects a change more quickly when the data exhibit underdispersion, that is, for a fixed value of d, the effectiveness of the new control chart increases as the value of the dispersion parameter increases. For example, when $d = 0.05$, the ARL_1 values for overdispersed ($\nu = 0.5$), equidispersed ($\nu = 1.0$), and underdispersed ($\nu = 2.0$) data are 149.55, 138.12, and 120.24, respectively. We also observe that the proposed chart detects upward shifts more quickly than downward shifts.

4.4.2 Steady-state ARL

The results provided in the previous section are obtained under the assumption that the shifts in the mean occur at the beginning of the process, that is, at time $t = 1$, which is known as zero-state ARL. However, in practice, this does not always happen as most times the process remains in a control condition for some time before the occurrence of the shift, which is known as the steady-state ARL (Lucas and Saccucci, 1990). Table 4.2 presents the steady-state ARL values of the proposed control chart for different values of $\nu \in \{0.5, 1.0, 2.0\}$. We assume that the shift occurs at different times $t \in \{1, 25, 50, 75, 100\}$, where $t = 1$ represents zero-state ARL. Note that, from the table, the proposed control chart becomes less effective at detecting shifts in the mean of the process as t increases, that is, the ability of the chart to detect shifts is reduced. This happens because the PM statistic is calculated using both current and past information.

Table 4.1: *Run length characteristics of the proposed chart when* $ARL_0 = 200$.

		d										
		−0.80	−0.40	−0.20	−0.10	−0.05	0.00	0.05	0.10	0.20	0.40	0.80
$\nu = 0.5$	ARL	17.89	35.88	73.94	129.53	175.78	199.85	149.55	97.86	52.97	22.63	8.77
	SDRL	10.41	27.03	74.42	168.02	262.40	290.24	236.62	144.37	67.39	24.53	8.18
	Q_1	2	2	2	2	2	2	2	2	1	1	1
	Q_{25}	10	16	21	21	22	29	16	14	10	6	2
	Q_{50}	16	30	51	71	74	86	58	44	29	14	6
	Q_{75}	24	49	103	170	215	241	174	125	70	31	12
	Q_{99}	48	122	343	791	1298	1461	1181	713	310	113	36
$\nu = 1.0$	ARL	15.67	28.74	60.47	112.40	169.78	200.96	138.12	84.10	41.21	17.06	6.55
	SDRL	8.83	20.27	57.87	140.04	251.08	302.23	214.58	119.59	48.28	17.12	5.73
	Q_1	2	2	2	2	2	2	2	1	1	1	1
	Q_{25}	9	14	19	22	22	30	16	13	9	4	2
	Q_{50}	14	25	44	63	73	85	56	42	24	12	5
	Q_{75}	21	39	83	151	208	232	162	109	56	24	9
	Q_{99}	41	93	276	645	1318	1530	1085	585	217	79	27
$\nu = 2.0$	ARL	12.96	22.04	46.09	91.40	150.01	200.01	120.24	70.36	31.25	12.31	4.51
	SDRL	7.08	14.40	40.68	104.93	212.67	299.63	182.72	92.81	33.77	11.64	3.69
	Q_1	2	2	2	2	2	2	1	1	1	1	1
	Q_{25}	8	11	17	20	21	28	14	12	8	3	2
	Q_{50}	12	19	35	56	72	86	50	36	20	9	3
	Q_{75}	17	30	64	125	191	232	144	95	43	17	6
	Q_{99}	32	65	187	474	1058	1589	868	452	155	53	17

Table 4.2: *Steady-state ARL of the proposed chart.*

		d										
	t	-0.80	-0.40	-0.20	-0.10	-0.05	0.00	0.05	0.10	0.20	0.40	0.80
$\nu = 0.5$	1	17.89	35.88	73.94	129.53	175.78	199.85	149.55	97.86	52.97	22.63	8.77
	25	21.96	40.65	80.65	141.27	192.84	205.98	169.47	117.51	63.69	29.48	12.59
	50	23.94	44.48	80.21	146.85	199.89	212.12	181.04	123.17	66.89	32.22	14.54
	75	25.90	46.53	84.74	151.54	204.18	216.62	176.45	125.83	69.42	33.77	15.47
	100	26.82	47.93	86.25	154.48	206.69	224.60	185.38	133.20	72.65	35.07	16.31
$\nu = 1.0$	1	15.67	28.74	60.47	112.40	169.78	200.96	138.12	84.10	41.21	17.06	6.55
	25	19.67	31.99	65.80	123.18	186.79	206.19	159.96	100.94	51.29	22.85	9.47
	50	21.50	33.46	67.80	126.79	190.22	211.60	165.54	106.55	55.47	25.63	11.20
	75	23.10	34.52	70.58	128.27	195.64	216.10	167.61	109.29	56.19	26.65	11.94
	100	24.16	35.72	71.70	129.37	197.91	223.75	175.11	115.79	58.66	27.85	12.77
$\nu = 2.0$	1	12.96	22.04	46.09	91.40	150.01	200.01	120.24	70.36	31.25	12.31	4.51
	25	16.65	24.75	51.18	98.54	166.17	205.78	138.76	84.51	38.94	17.08	6.68
	50	18.66	27.18	53.80	102.58	169.05	211.04	150.64	87.20	42.06	19.22	8.09
	75	19.86	29.04	54.97	104.45	176.14	218.29	153.00	89.76	44.33	20.60	8.93
	100	21.37	30.23	56.85	106.98	179.86	224.62	157.92	92.30	45.52	21.26	9.43

Table 4.3: *Design structures of the control charts.*

Control chart	Plotting statistic	Control limits
Shewhart-type	$r_t^{(q)}$	$UCL_t = \mu_0 + L\sqrt{\sigma_0^2}$ $LCL_t = \mu_0 - L\sqrt{\sigma_0^2}$
EWMA	$W_t = (1-\lambda)W_{t-1} + \lambda r_t^{(q)}$	$UCL_t = \mu_0 - L\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2t}]\sigma_0^2}$ $UCL_t = \mu_0 + L\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2t}]\sigma_0^2}$
Proposed	$PM_t = \frac{\sum_{i=1}^{t} r_i^{(q)}}{t}$	$UCL_t = \mu_0 - L\frac{1}{f(t)}\sqrt{\sigma_0^2}$ $UCL_t = \mu_0 + L\frac{1}{f(t)}\sqrt{\sigma_0^2}$

4.5 Comparative study

In this section, we analyze and compare the performance of the proposed control chart with the traditional Shewhart and EWMA control charts. Table 4.3 summarizes the design structures of these charts. For the EWMA chart, we considered the smoothing parameter value $\lambda = 0.20$.

The performance of the control charts is measured in terms of out-of-control average run lengths (ARL_1). For that, the value of the width of the control limit (L) is set to satisfy an ARL_0 equal or approximately equal to a fixed value (here $\text{ARL}_0 = 200$) and then the ARL_1 values are calculated and compared. When the process is in control, a large ARL_0 is desirable. Alternatively, when the process is out of control, we desire the lowest ARL_1.

Table 4.4 presents the results for the performance comparison of the three control charts when $\text{ARL}_0 = 200$. As becomes clear from this table, the proposed chart outperforms the other two charts in all scenarios and shifts considered in the process. We also observe that the Shewhart and EWMA charts are biased for small downward shifts, that is, the ARL_1 values are greater than the corresponding ARL_0 values. For example, when $\nu = 2.0$ and $d = -0.05$, the ARL_1 values for the Shewhart and EWMA charts are 213.79 and 224.33, respectively, while the proposed control chart presents an ARL_1 value of 150.01.

4.6 Illustrative example

In this section, we illustrate the applicability of the proposed control chart using a real data example. We consider the dataset of the weekly hospital admissions due to respiratory diseases of people aged over 60 years in the city of São Paulo–Brazil from January 2010 to December 2015. The dataset was obtained from the Hospitalization Information System of the Ministry of Health (available at Datasus website `http://datasus.saude.gov.br/`) and previously analyzed in Melo and Alencar (2020). São Paulo is the most populous city of Brazil with 11 million inhabitants in 2010, 11% of whom are over 60 years old (`http://www.ibge.gov.br/`). Given its relevance, the monitoring and identification of patterns in the weekly number of hospitalizations due to respiratory problems may be useful to plan health services. For example, based on these results, governments may create vaccination strategies (Alencar, 2018).

In phase I, we considered the weekly data from January 2010 to December 2014 to estimate the

Table 4.4: *ARL comparison between the Shewhart, EWMA, and proposed charts for different shifts when* $ARL_0 = 200$.

Control charts	d	Dispersion		
		$\nu = 0.5$	$\nu = 1.0$	$\nu = 2.0$
Shewhart-type	−0.80	234.12	160.21	53.42
	−0.40	252.09	212.16	137.99
	−0.20	244.72	229.70	206.86
	−0.10	224.14	222.79	215.05
	−0.05	210.28	210.38	213.79
	0.00	197.15	197.94	199.18
	0.05	177.76	180.66	181.32
	0.10	162.90	158.97	160.63
	0.20	130.81	123.08	117.09
	0.40	74.59	64.92	51.90
	0.80	26.42	18.70	11.60
		$L = 2.807$	$L = 2.807$	$L = 2.807$
PM statistic	−0.80	17.89	15.67	12.96
	−0.40	35.88	28.74	22.04
	−0.20	73.94	60.47	46.09
	−0.10	129.53	112.40	91.40
	−0.05	175.78	169.78	150.01
	0.00	199.85	200.96	200.01
	0.05	149.55	138.12	120.24
	0.10	97.86	84.10	70.36
	0.20	52.97	41.21	31.25
	0.40	22.63	17.06	12.31
	0.80	8.77	6.55	4.51
		$L = 3.243$	$L = 3.244$	$L = 3.244$
EWMA	−0.80	66.04	53.44	36.37
	−0.40	136.9	111.44	78.76
	−0.20	221.16	193.78	155.98
	−0.10	237.57	230.22	213.06
	−0.05	225.52	232.11	224.33
	0.00	200.07	200.27	200.48
	0.05	170.08	162.94	154.21
	0.10	134.96	125.16	111.05
	0.20	82.38	67.98	53.44
	0.40	33.38	25.02	17.46
	0.80	10.48	7.36	4.88
		$L = 2.655$	$L = 2.655$	$L = 2.655$

Table 4.5: *Fitted CMP-ARMA(2, 0) model for the weekly number of hospitalizations data.*

Coefficient	Estimate	SE	p-value
intercept (α)	2.6812	0.3760	< 0.0001
Cosine (β_1)	-0.1022	0.0180	< 0.0001
ϕ_1	0.3504	0.0631	< 0.0001
ϕ_2	0.1693	0.0629	0.0071
dispersion (ν)	0.3986	0.0327	< 0.0001

model parameters and determine the control limits of the charts. As in Melo and Alencar (2020), we consider a CMP-ARMA$(2, 0)$ model with mean

$$\log(\mu_t) = \alpha + \beta_1 \cos\left(\frac{2\pi t}{52}\right) + \phi_1 \log(y_{t-1}) + \phi_2 \log(y_{t-2}),$$

where t is the number of each week, $t = 1, \ldots, 262$. Table 4.5 shows the parameter estimates with corresponding standard errors (SE) and p-values for the fitted model. Note that the estimated dispersion parameter was $\nu = 0.3986$, showing that the data are overdispersed. Based on the estimated parameters, we obtain L to meet an $ARL_0 = 200$ using the simulation algorithm introduced in Subsection 4.4.1. The width of the control limits for the Shewhart-type, PM, and EWMA ($\lambda = 0.2$) control charts are, respectively, $L = 2.807$, $L = 3.243$, and $L = 2.655$.

In phase II, we evaluate the efficiency of the proposed chart by monitoring a series consisting of 52 observations. The first half of the sample contains the first 26 observations of the weekly number of hospitalizations in 2015, thus these observations are assumed to be in control. The remaining of the series is simulated from a CMP-ARMA$(2, 0)$ model with out-of-control mean $\mu_{1,t} = 1.1 \exp\{\widehat{\alpha} + \widehat{\beta}_1 \cos(2\pi t/52) + \widehat{\phi}_1 \log(y_{t-1}) + \widehat{\phi}_2 \log(y_{t-2})\}$, $t = 27, \ldots, 52$, and dispersion parameter $\widehat{\nu}$, where $(\widehat{\alpha}, \widehat{\beta}, \widehat{\phi}_1, \widehat{\phi}_2, \widehat{\nu})$ are the estimates shown in Table 4.5. That is, we assume that after the 26th observation we have an increase of 10% in the process mean.

Fugures 4.1, 4.2, and 4.3 display the three control charts based on the randomized quantile residuals obtained from the fitted model. The red dots correspond to observations that lie outside the control limits. These figures show that in the three charts all the first 26 points fall inside the control limits, showing that the process is in control. Note that even when we have an out-of-control process situation, after the 26th observation, the Shewhart chart cannot detect the shift, while the PM and EWMA control charts give the first out-of-control signal at the 31th and 32th observations, respectively. Also, note that, in addition to detecting the out-of-control process faster, the proposed chart provides more out-of-control points compared to the other two charts considered in this article.

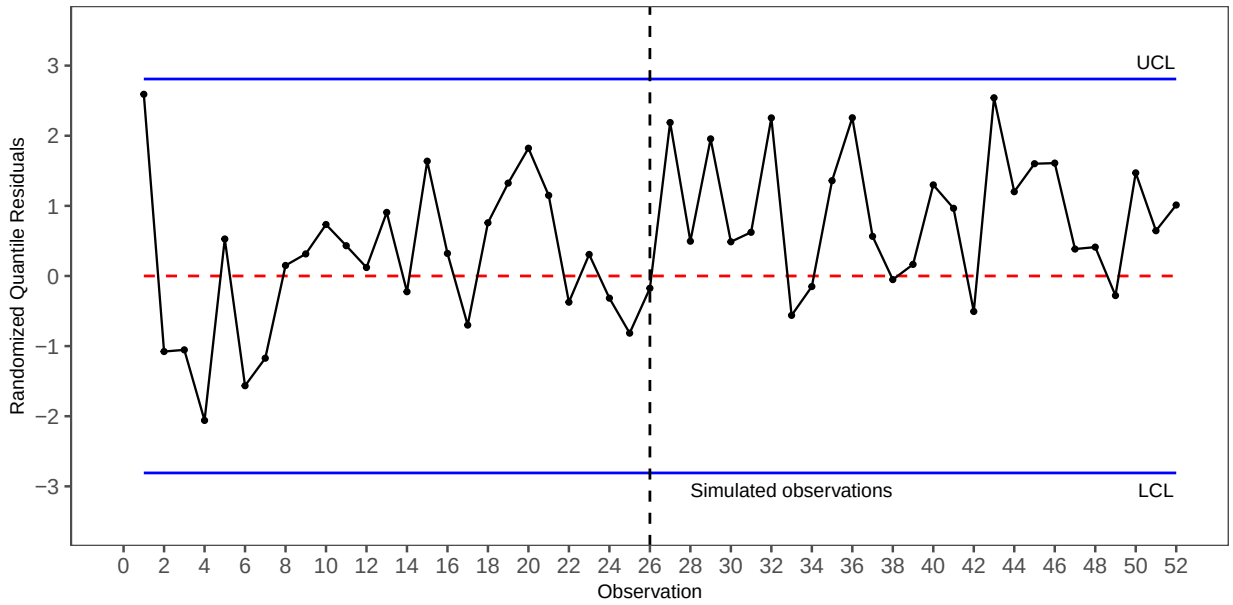


Figure 4.1: *The Shewart-type chart for the illustrative example.*

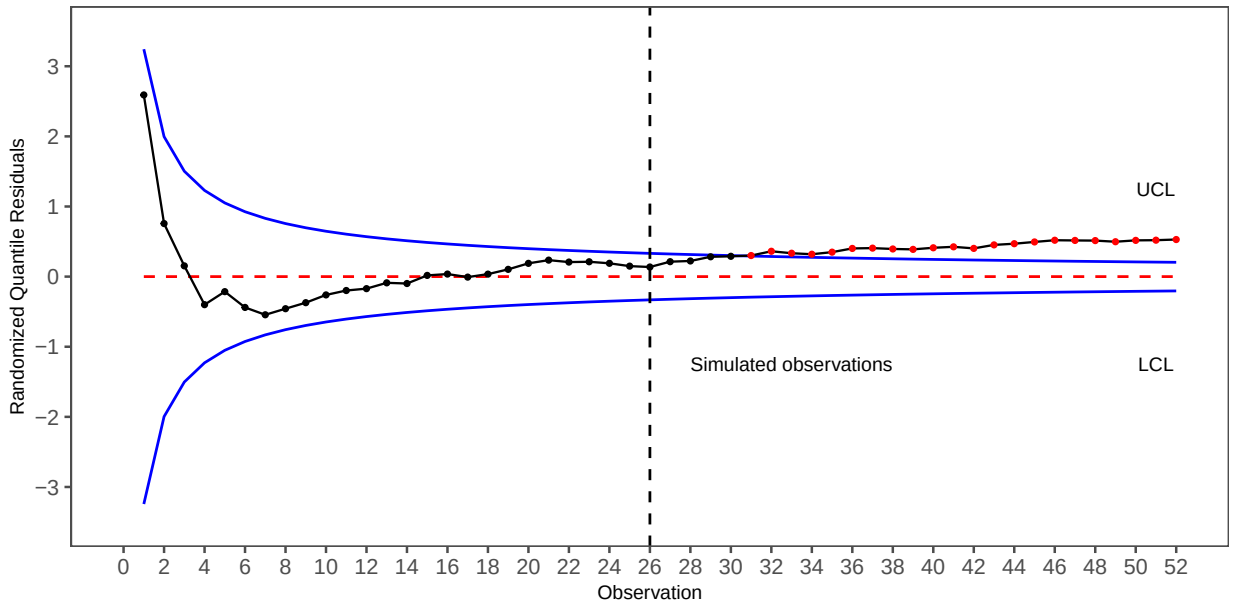


Figure 4.2: *The PM chart (proposed) for the illustrative example.*

4.7 Conclusions

In this article, we proposed a new memory-type control chart to detect a shift in the mean of count times series. The new monitoring procedure is based on randomized quantile residuals obtained from a fitted CMP-ARMA model. The proposed control chart can be used to monitor underdispersed, equidispersed, and overdispersed time series count data. We conducted a Monte Carlo simulation study to evaluate the performance of the proposed control chart. Numerical results showed that the current proposal presents a good performance and can detect shifts faster than the EWMA and Shewhart-type control charts in terms of average run lengths. We also investigated

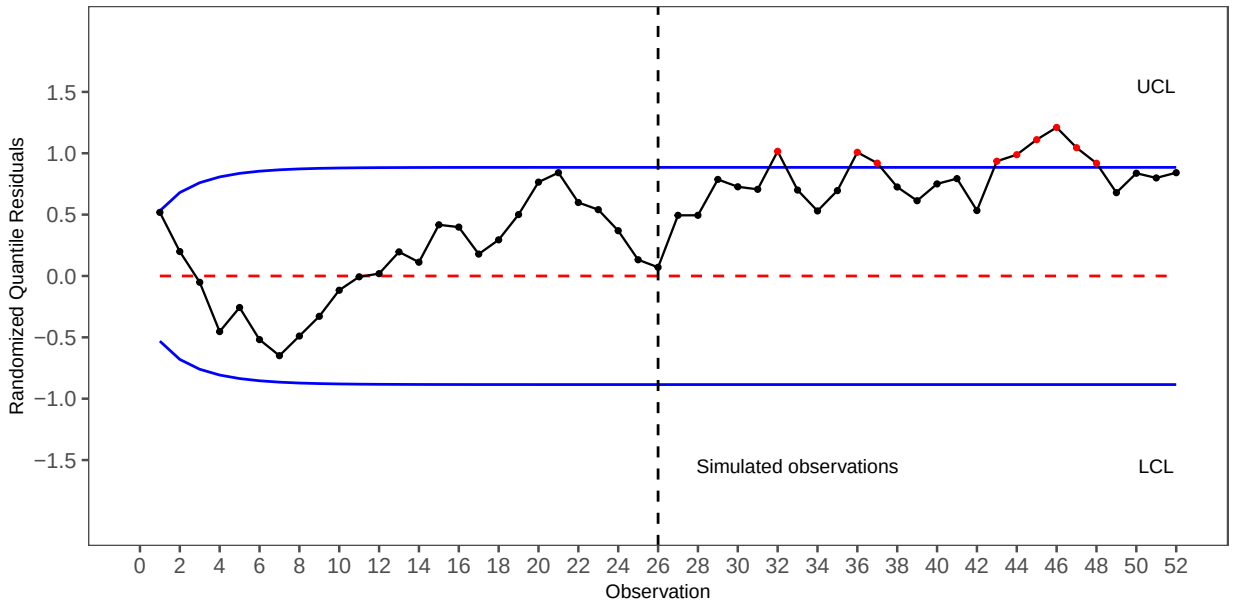


Figure 4.3: *The EWMA chart for the illustrative example.*

the steady-state ARL performance of the proposed chart assuming that a shift in the process mean occurs at different times t. We observed that the chart detects the out-of-control condition in the process more slowly as t increases. Finally, we illustrated the applicability of the proposed control chart by monitoring the weekly number of hospitalizations due to respiratory diseases of people older than 60 years in São Paulo city, where the new chart detected more quickly the shift in the process mean.

Chapter 5

Conclusions

In this thesis, we proposed two new dynamic regression models based on the Conway-Maxwell-Poisson distribution for the analysis of time series of counts. In addition, we introduced a new memory-type control chart for monitoring this type of data.

The first proposed model, named the Conway-Maxwell-Poisson autoregressive moving average model (CMP-ARMA), is a dynamic model for random variables that assume non-negative integer values. In the second model proposed, we extend the class of Conway-Maxwell-Poisson autoregressive moving average models by including seasonal components to the model dynamic structure. In the construction of the models, we assume that the conditional distribution of the dependent variable, given the past history of the process, is the in CMP (Huang, 2017). This distribution allows the modeling of underdispersed, equidispersed, and overdispersed data. As in Benjamin *et al.* (2003), the conditional mean of the distribution is modeled by a dynamic structure containing autoregressive and moving average terms, time-varying regressors, and a link function. For the two models proposed, we discuss methods for parameter estimation, hypothesis testing inference, and diagnostic analysis; obtaining closed-form expressions for the score vector and Fisher information matrix. The results of the Monte Carlo simulation study carried out showed that the conditional maximum likelihood estimators presents a good performance, suggesting that the estimators of the parameters are unbiased and consistent. The empirical applications show that the proposed models better accommodated the observations with subdispersion, when compared to the results provided by the negative binomial GARMA model. When the observations show overdispersion, the models show similar results.

Finally, we propose a new memory-type control chart based on the progressive mean statistic for monitoring autocorrelated count. The new is based on randomized quantile residuals obtained from a fitted Conway-Maxwell-Poisson autoregressive moving average model. A simulation study is carried out to evaluate the performance of the proposed control chart. The results revealed that the proposed chart outperformed the tradicional Shewhart and EWMA control charts in terms of the average run length.

5.1 Suggestions for Future Research

- To develop and study a new class of models for Comway-Maxwell-Poisson conditional distribution with heteroscedasticity.

- To investigate bias correction methods in small samples of the conditional maximum likelihood estimators of both proposed models CMP-ARMA and CMP-SARMA.

Bibliography

Abbas(2015) Nasir Abbas. Progressive mean as a special case of exponentially weighted moving average. *Quality and Reliability Engineering International*, 31(4):719–720. Page. 52

Abbas *et al.***(2011)** Nasir Abbas, Muhammad Riaz and Ronald JMM Does. Enhancing the performance of EWMA charts. *Quality and Reliability Engineering International*, 27(6):821–833. Page. 52

Abbas *et al.***(2013)** Nasir Abbas, Raja Fawad Zafar, Muhammad Riaz and Zawar Hussain. Progressive mean control chart for monitoring process location parameter. *Quality and Reliability Engineering International*, 29(3):357–367. Page. 52, 54, 55

Abbasi(2017) Saddam Akber Abbasi. Poisson progressive mean control chart. *Quality and Reliability Engineering International*, 33(8):1855–1859. Page. 52, 55

Abbasi and Miller(2013) Saddam Akber Abbasi and Arden Miller. Mdewma chart: an efficient and robust alternative to monitor process dispersion. *Journal of Statistical Computation and Simulation*, 83(2):247–268. Page. 52

Abbasi *et al.***(2012)** Saddam Akber Abbasi, Muhammad Riaz and Arden Miller. Enhancing the performance of CUSUM scale chart. *Computers & Industrial Engineering*, 63(2):400–409. Page. 52

Abbasi *et al.***(2013)** Saddam Akber Abbasi, Arden Miller and Muhammad Riaz. Nonparametric progressive mean control chart for monitoring process target. *Quality and Reliability Engineering International*, 29(7):1069–1080. Page. 52, 55

Abbasi *et al.***(2019)** Saddam Akber Abbasi, Mohamed Abbas Mohamed, Mohamed Adil Ahmed, Rommel Joseph Lajara and Hadi Fadel Hadi. Monitoring coefficient of variation using progressive mean technique. In *2019 8th International Conference on Industrial Technology and Management (ICITM)*, pp. 270–274. IEEE. Page. 52

Akaike(1974) Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723. Page. 13, 42

Albarracin *et al.***(2018a)** Orlando Yesid Esparza Albarracin, Airlane Pereira Alencar and Linda Lee Ho. Effect of neglecting autocorrelation in regression EWMA charts for monitoring count time series. *Quality and Reliability Engineering International*, 34(8):1752–1762. Page. 35, 52

Albarracin *et al.***(2018b)** Orlando Yesid Esparza Albarracin, Airlane Pereira Alencar and Linda Lee Ho. CUSUM chart to monitor autocorrelated counts using negative binomial GARMA model. *Statistical methods in medical research*, 27(9):2859–2871. Page. 1, 35, 51

Albarracin *et al.***(2019)** Orlando Yesid Esparza Albarracin, Airlane Pereira Alencar and Linda Lee Ho. Generalized autoregressive and moving average models: multicollinearity, interpretation and a new modified model. *Journal of Statistical Computation and Simulation*, 89(10):1819–1840. Page. 25

Alencar(2018) Airlane P Alencar. Seasonality of hospitalizations due to respiratory diseases: modelling serial correlation all we need is Poisson. *Journal of Applied Statistics*, 45(10):1813–1822. Page. 17, 59

Alevizakos and Koukouvinos(2019a) Vasileios Alevizakos and Christos Koukouvinos. A double exponentially weighted moving average control chart for monitoring COM-Poisson attributes. *Quality and Reliability Engineering International*, 35(7):2130–2151. Page. 2, 52

Alevizakos and Koukouvinos(2019b) Vasileios Alevizakos and Christos Koukouvinos. A progressive mean control chart for COM-Poisson distribution. *Communications in Statistics-Simulation and Computation*, pp. 1–19. Page. 2, 52, 55

Andersen(1970) Erling Bernhard Andersen. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 283–301. Page. 11, 42

Aslam *et al.***(2016)** Muhammad Aslam, Liaquat Ahmad, Chi-Hyuck Jun and Osama H Arif. A control chart for com–poisson distribution using multiple dependent state sampling. *Quality and Reliability Engineering International*, 32(8):2803–2812. Page. 1, 52

Aslam *et al.***(2017)** Muhammad Aslam, Aamir Saghir, Liaquat Ahmad, Chi-Hyuck Jun and Jaffer Hussain. A control chart for COM-Poisson distribution using a modified EWMA statistic. *Journal of Statistical Computation and Simulation*, 87(18):3491–3502. Page. 2, 52

Aslam *et al.***(2018)** Muhammad Aslam, Nasrullah Khan and Chi-Hyuck Jun. A hybrid exponentially weighted moving average chart for COM-Poisson distribution. *Transactions of the Institute of Measurement and Control*, 40(2):456–461. Page. 2, 52

Bayer *et al.***(2018)** Fábio M Bayer, Renato J Cintra and Francisco Cribari-Neto. Beta seasonal autoregressive moving average models. *Journal of Statistical Computation and Simulation*, 88 (15):2961–2981. Page. 36, 37, 42, 44

Bayer *et al.***(2017)** Fábio Mariano Bayer, Débora Missio Bayer and Guilherme Pumi. Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology*, 555:385–396. Page. 6, 9, 36, 42

Benjamin *et al.***(1998)** Michael A Benjamin, Robert A Rigby and Mikis D Stasinopoulos. Fitting non-gaussian time series models. In *COMPSTAT Proceedings in Computational Statistics Heidelburg: Physica-Verlag*, pp. 191–196. Page. 11

Benjamin *et al.***(2003)** Michael A Benjamin, Robert A Rigby and D Mikis Stasinopoulos. Generalized autoregressive moving average models. *Journal of the American Statistical association*, 98 (461):214–223. Page. 1, 5, 6, 9, 10, 13, 15, 23, 35, 43, 52, 53, 65

Bourguignon *et al.***(2016)** Marcelo Bourguignon, Klaus LP Vasconcellos, Valdério A Reisen and Márton Ispány. A poisson INAR (1) process with a seasonal structure. *Journal of Statistical Computation and Simulation*, 86(2):373–387. Page. 36, 44

Box *et al.***(2015)** George EP Box, Gwilym M Jenkins, Gregory C Reinsel and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons. Page. 1, 5, 36

Briët *et al.***(2013)** Olivier JT Briët, Priyanie H Amerasinghe and Penelope Vounatsou. Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. *PLoS One*, 8(6):e65761. Page. 36

Capizzi and Masarotto(2003) Giovanna Capizzi and Guido Masarotto. An adaptive exponentially weighted moving average control chart. *Technometrics*, 45(3):199–207. Page. 52

Conway and Maxwell(1962) Richard W Conway and William L Maxwell. A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12(2):132–136. Page. 2, 7

Czado *et al.***(2009)** Claudia Czado, Tilmann Gneiting and Leonhard Held. Predictive model assessment for count data. *Biometrics*, 65(4):1254–1261. Page. 14, 43

Dugas *et al.***(2013)** Andrea Freyer Dugas, Mehdi Jalalpour, Yulia Gel, Scott Levin, Fred Torcaso, Takeru Igusa and Richard E Rothman. Influenza forecasting with google flu trends. *PloS one*, 8 (2). Page. 35

Dunn and Smyth(1996) Peter K Dunn and Gordon K Smyth. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244. Page. 13, 43, 52

Evans(1953) DA Evans. Experimental evidence concerning contagious distributions in ecology. *Biometrika*, 40(1/2):186–211. Page. 17, 44

Feng *et al.***(2017)** Cindy Feng, Alireza Sadeghpour and Longhai Li. Randomized quantile residuals: an omnibus model diagnostic tool with unified reference distribution. *arXiv preprint arXiv:1708.08527*. Page. 13, 43, 54

Fokianos and Kedem(2004) Konstantinos Fokianos and Benjamin Kedem. Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis*, 25 (2):173–197. Page. 6, 13, 43

Franke and Seligmann(1993) J Franke and Th Seligmann. Conditional maximum likelihood estimates for INAR (1) processes and their application to modelling epileptic seizure counts. *Developments in time series*, 310:330. Page. 5

Freeland(1998) R Keith Freeland. *Statistical analysis of discrete time series with application to the analysis of workers' compensation claims data*. Tese de Doutorado, University of British Columbia. Page. 44

Freeland and McCabe(2004) RK Freeland and Brendan PM McCabe. Analysis of low count time series data by poisson autoregression. *Journal of Time Series Analysis*, 25(5):701–722. Page. 5

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

Gay(1990) David M Gay. Usage summary for selected optimization routines. *Computing science technical report*, 153:1–21. Page. 11, 39

Haq *et al.***(2014)** Abdul Haq, Jennifer Brown and Elena Moltchanova. A new cumulative sum quality control scheme for monitoring the process mean. *Quality and Reliability Engineering International*, 30(8):1165–1177. Page. 52

Haq *et al.***(2015)** Abdul Haq, Jennifer Brown and Elena Moltchanova. New exponentially weighted moving average control charts for monitoring process mean and process dispersion. *Quality and Reliability Engineering International*, 31(5):877–901. Page. 52

Huang(2017) Alan Huang. Mean-parametrized Conway–Maxwell–Poisson regression models for dispersed counts. *Statistical Modelling*, 17(6):359–380. Page. 2, 6, 8, 9, 23, 27, 36, 47, 53, 65

Jiang *et al.***(2008)** Wei Jiang, Lianjie Shu and Daniel W Apley. Adaptive CUSUM procedures with EWMA-based shift estimators. *Iie Transactions*, 40(10):992–1003. Page. 52

Jung and Tremayne(2006) Robert C Jung and AR Tremayne. Binomial thinning models for integer time series. *Statistical Modelling*, 6(2):81–96. Page. 19

Jung and Tremayne(2011) Robert C Jung and AR Tremayne. Useful models for time series of counts or simply wrong ones? *AStA Advances in Statistical Analysis*, 95(1):59–91. Page. 14, 43

Jung *et al.***(2016)** Robert C Jung, Brendan PM McCabe and Andrew R Tremayne. Model validation and diagnostics. In *Handbook of discrete-valued time series*, pp. 189–218. Boca Raton, FL: Chapman & Hall/CRC. Page. 14, 43

Kedem and Fokianos(2005) Benjamin Kedem and Konstantinos Fokianos. *Regression models for time series analysis*, volume 488. John Wiley & Sons. Page. 12, 13, 42

Li(1991) WK Li. Testing model adequacy for some markov regression models for time series. *Biometrika*, 78(1):83–89. Page. 6

Li(1994) WK Li. Time series models based on generalized linear models: some further results. *Biometrics*, pp. 506–511. Page. 6

Liboschik *et al.***(2017)** Tobias Liboschik, Konstantinos Fokianos and Roland Fried. tscount: An R package for analysis of count time series following generalized linear models. *Journal of Statistical Software*, 82(5):1–51. Page. 6

Ljung and Box(1978) Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303. Page. 18, 47

Lord *et al.*(2008) Dominique Lord, Seth D Guikema and Srinivas Reddy Geedipally. Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis & Prevention*, 40(3):1123–1134. Page. 2

Lord *et al.*(2012) Dominique Lord, Seth D Guikema *et al.* The Conway–Maxwell–Poisson model for analyzing crash data. *Applied Stochastic Models in Business and Industry*, 28(2):122–127. Page. 2

Lucas and Saccucci(1990) James M Lucas and Michael S Saccucci. Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, 32(1):1–12. Page. 56

MacDonald and Bhamani(2018) Iain L MacDonald and Feroz Bhamani. A time-series model for underdispersed or overdispersed counts. *The American Statistician*, pp. 1–12. Page. 6, 19

Maior and Cysneiros(2018) Vinicius QS Maior and Francisco José A Cysneiros. SYMARMA: a new dynamic model for temporal data on conditional symmetric distribution. *Statistical Papers*, 59(1):75–97. Page. 36

Mamode Khan *et al.*(2018) Naushad Mamode Khan, Vandna Jowaheer, Yuvraj Sunecher and Marcelo Bourguignon. Modeling longitudinal INMA (1) with COM–Poisson innovation under non-stationarity: application to medical data. *Computational and Applied Mathematics*, pp. 1–22. Page. 1, 6

McCullagh and Nelder(1989) P. McCullagh and J.A. Nelder. *Generalized linear models*. Chapman and Hall, 2nd edição. Page. 6, 35

McCullagh(1984) Peter McCullagh. Generalized linear models. *European Journal of Operational Research*, 16(3):285–292. Page. 2, 3

Melo and Alencar(2020) Moizes Melo and Airlane Alencar. Conway-maxwell-poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data. *Journal of Time Series Analysis*. Page. 36, 37, 40, 53, 54, 59, 61

Monteiro *et al.*(2010) Magda Monteiro, Manuel G Scotto and Isabel Pereira. Integer-valued autoregressive processes with periodic structure. *Journal of Statistical Planning and Inference*, 140(6):1529–1541. Page. 36

Montgomery(2007) Douglas C Montgomery. *Introduction to statistical quality control*. John Wiley & Sons. Page. 51

Page(1954) Ewan S Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115. Page. 52

Park *et al.*(2020) Kayoung Park, Dongmin Jung and Jong-Min Kim. Control charts based on randomized quantile residuals. *Applied Stochastic Models in Business and Industry*. Page. 52

Pawitan(2001) Yudi Pawitan. *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press. Page. 12, 42

Pumi *et al.*(2019) Guilherme Pumi, Marcio Valk, Cleber Bisognin, Fábio Mariano Bayer and Taiane Schaedler Prass. Beta autoregressive fractionally integrated moving average models. *Journal of Statistical Planning and Inference*, 200:196–212. Page. 42

R Core Team(2019) R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/. Page. 11, 15, 39, 44, 55

Rao *et al.*(2020) Gadde Srinivasa Rao, Muhammad Aslam, Umer Rasheed and Chi-Hyuck Jun. Mixed EWMA–CUSUM chart for COM-Poisson distribution. *Journal of Statistics and Management Systems*, 23(3):511–527. Page. 2, 52

Roberts(1959) SW Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 1(3):239–250. Page. 52

Rocha and Cribari-Neto(2009) Andréa V Rocha and Francisco Cribari-Neto. Beta autoregressive moving average models. *Test*, 18(3):529–545. Page. 6, 9, 35

Saghir and Lin(2014a) Aamir Saghir and Zhengyan Lin. Control chart for monitoring multivariate COM-Poisson attributes. *Journal of Applied Statistics*, 41(1):200–214. Page. 1, 52

Saghir and Lin(2014b) Aamir Saghir and Zhengyan Lin. Cumulative sum charts for monitoring the COM-Poisson processes. *Computers & Industrial Engineering*, 68:65–77. Page. 1, 52

Saghir and Lin(2014c) Aamir Saghir and Zhengyan Lin. A flexible and generalized exponentially weighted moving average control chart for count data. *Quality and Reliability Engineering International*, 30(8):1427–1443. Page. 51

Schwarz *et al.*(1978) Gideon Schwarz *et al.* Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464. Page. 13, 43

Sellers(2012) Kimberly F Sellers. A generalized statistical control chart for over-or under-dispersed data. *Quality and Reliability Engineering International*, 28(1):59–65. Page. 1, 52

Sellers and Morris(2017) Kimberly F Sellers and Darcy S Morris. Underdispersion models: Models that are “under the radar". *Communications in Statistics-Theory and Methods*, 46(24): 12075–12086. Page. 1

Sellers and Shmueli(2010) Kimberly F Sellers and Galit Shmueli. A flexible regression model for count data. *The Annals of Applied Statistics*, pp. 943–961. Page. 7

Sellers *et al.*(2012) Kimberly F Sellers, Sharad Borle and Galit Shmueli. The COM-Poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, 28(2):104–116. Page. 1, 2, 6

Shewhart(1927) WA Shewhart. Quality control. *The Bell System Technical Journal*, 6(4):722–735. Page. 51

Shewhart(1926) Walter A Shewhart. Quality control charts. *The Bell System Technical Journal*, 5(4):593–603. Page. 51

Shmueli *et al.***(2005)** Galit Shmueli, Thomas P Minka, Joseph B Kadane, Sharad Borle and Peter Boatwright. 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. Page. 2, 6, 7, 8

Shu and Jiang(2008) Lianjie Shu and Wei Jiang. A new EWMA chart for monitoring process dispersion. *Journal of Quality Technology*, 40(3):319–331. Page. 52

Stasinopoulos and Rigby(2016) Maintainer Mikis Stasinopoulos and Bob Rigby. Package gamlss.util. *page*, 9. Page. 17, 46

Sunecher *et al.***(2018)** Yuvraj Sunecher, Naushad Mamode Khan and Vandna Jowaheer. BINMA (1) model with COM-Poisson innovations: Estimation and application. *Communications in Statistics-Simulation and Computation*, 49(6):1631–1652. Page. 1

Talamantes *et al.***(2007)** Jorge Talamantes, Sam Behseta and Charles S Zender. Statistical modeling of valley fever data in kern county, california. *International journal of biometeorology*, 51 (4):307. Page. 35

Tiku *et al.***(2000)** Moti L Tiku, Wing-Keung Wong, David C Vaughan and Guorui Bian. Time series models in non-normal situations: Symmetric innovations. *Journal of Time Series Analysis*, 21(5):571–596. Page. 36

Zeger and Qaqish(1988) Scott L Zeger and Bahjat Qaqish. Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, pp. 1019–1031. Page. 5

Zhu(2012) Fukang Zhu. Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling*, 56(9-10):191–203. Page. 1, 6