



**HAL**  
open science

# Robust estimate for count time series using GLARMA models: An application to environmental and epidemiological data

Ana Julia Alves Camara, Valderio Anselmo Reisen, Pascal Bondon

## ► To cite this version:

Ana Julia Alves Camara, Valderio Anselmo Reisen, Pascal Bondon. Robust estimate for count time series using GLARMA models: An application to environmental and epidemiological data. *Applied Mathematical Modelling*, 2025, 137 (Part A), pp.115658. 10.1016/j.apm.2024.115658 . hal-04691875

**HAL Id: hal-04691875**

**<https://hal.science/hal-04691875>**

Submitted on 9 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Robust estimate for count time series using GLARMA models: An application to environmental and epidemiological data

Ana Julia Alves Camara<sup>1</sup>, Valderio Anselmo Reisen<sup>2</sup> and Pascal Bondon<sup>3</sup>

<sup>1</sup>Federal University of Minas Gerais, Department of Statistics, Brazil.

<sup>2</sup>Federal University of Espirito Santo, Department of Statistics, Brazil.

<sup>3</sup>Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France.

## Abstract

The Generalized Linear Autoregressive Moving Average (GLARMA) model has been used in epidemiological studies to evaluate the impact of air pollutants on health. Due to the nature of the data, a robust approach for the GLARMA model is proposed here based on the robustification of the quasi-likelihood function. Outlying observations are bounded separately by weight functions on covariates and the Huber loss function on the response variable. Some technical issues related to the robust approach are discussed and a Monte Carlo study revealed that the robust approach is more reliable than the classic one for contaminated data with additive outliers. The real data analysis investigates the impact of PM<sub>10</sub> in the number of deaths by respiratory diseases in Vitória, Brazil.

**Keywords.** Count time series, GLARMA model, M-estimators, Additive outliers, Respiratory diseases

**MSC 2020.** Primary : 62J12, 62F35 ; Secondary : 62M10, 62P12

## 1 Introduction

The expansion of cities and communities in the last decades led to economic growth and urban development. However, it also originated environmental and health problems once many activities generate residues that affect the population quality of life. Ozone (O<sub>3</sub>), nitrogen dioxide (NO<sub>2</sub>), sulfur dioxide (SO<sub>2</sub>), carbon monoxide (CO), and particulate matter (PM) are the main pollutants in the atmosphere, and even at concentrations within limits established by the World Health Organization (WHO) present a risk to human health (Pope and Dockery (2018), and Lippmann (2014)). Epidemiological studies have shown evidence of an association between concentration levels of air pollutants and mortality, morbidity, and hospital admissions, mainly caused by respiratory and cardiovascular diseases, see Pope et al. (1995); Dockery and Pope (1996); Ostro et al. (1999); Schwartz (2000); Ostro et al. (2009); Chen et al. (2010), and Froes et al. (2016), among others.

Epidemiological data are frequently treated as count time series as they record the frequency of events in successive time intervals. Count series are non-Gaussian processes formed by non-negative integers. They naturally arise in scientific areas such as economy, medicine, agriculture, and sports, among others. Examples are the monthly number of hospital admissions caused by a disease, car

accidents in a city, and transactions of a given stock observed in one hour. Methodologies started to emerge in the early 1970s. Initially, count time series were adjusted by Generalized Linear Models (GLM), introduced by Nelder and Wedderburn (1972). This procedure expands to the exponential family the possible distribution of the response variable, e.g. Normal, Poisson, Gamma, Negative Binomial, etc. In addition, the relation between the mean of the dependent variable ( $\mu$ ) and the linear predictor ( $\eta$ ) can be more flexible, assuming any monotonous non-linear function. Nevertheless, the GLM can not capture the time dependency structure in the data. The earliest work considering correlated time series can be found in Cox (1981), where models are classified into two categories: observation and parameter-driven. The main difference between them is how the dependence structure is added to the model. Zeger and Qaqish (1988) proposed a quasi-likelihood approach to time series regression, generalized by Benjamin et al. (2003). Davis et al. (1999) and Davis et al. (2003) introduced the GLARMA model, which consists in adding an Autoregressive Moving Average (ARMA) structure (Box and Jenkins, 1976) to the Generalized Linear (GLM) Model, allowing the modeling of correlated observations from the exponential family. Fokianos and Tjostheim (2011) proposed log-linear models for time series. Other procedures can be found in the recent overview of methodologies for count time series realized by Davis et al. (2021). Although many methods have been developed, they all present limitations that contribute to the non-development of a unified theory. Despite this, Davis et al. (2021) argue that the GLARMA family is “one of the most flexible and easily fit count models that balance parameter and observation-driven models”. Even though the GLARMA model presents some limitations regarding properties for general models, this method has been widely used in applications in distinct fields of knowledge; see, e.g. Rydberg and Shephard (2003) in finances, Karami et al. (2017) in air pollution, Kim et al. (2018) in engineering, Ballesteros-Cánovas et al. (2018) and Peitzsch et al. (2021) in climate changes, among others.

The statistical association between air quality variables and health effects has to consider some factors independently of the methodology applied or the type of response variable. In the epidemiological context, the response variable is usually time-correlated, a data property that must be considered in the step modeling. Furthermore, as the population under study is exposed to the complex mixture of pollutants and chemical compounds, the dynamics of the response variable and, therefore, the statistical functions that measure the impact of pollutants on health cannot be fully explained by the response variable itself or by just one contaminant.

Many authors have ignored the fact that the contaminants present multicollinearity. Souza et al. (2018) showed that if this characteristic is not treated properly, the association measures can be significantly impacted, leading to false conclusions regarding the population’s health risk in generalized additive models.

Here, the covariates are time-correlated and display complex behaviors such as periodicity, missing values, and extreme observations. High levels, or peaks, of pollutants, are frequently observed in air quality variables and often ignored. However, they can affect the estimation of some data characteristics, like the sample mean, variance, and correlation. In addition, many authors have verified that atypical observations (outliers) can seriously deteriorate the estimates of time series models (Reisen et al. (2017)).

Robustness indicates insensitivity to minor deviations from the assumptions (Huber (1981)). The foundations of this statistical approach can be found in Tukey (1960), Huber (1964), and Hampel (1968). Robust models have the characteristic of fitting properly to most datasets. If the data has no abrupt observations, the robust method will behave approximately the same as the classic model. Nevertheless, if the data has a small percentage of outliers, the robust models will show results almost as good as the classic models applied to clean data. Usually, robust estimates depend on a dispersion function that varies more slowly in extreme values than the quadratic functions. Outliers in time series can seriously affect the estimation and inference of parameters (Martin and Yohai (1985) and Bustos and Yohai (1986)). Fox (1972) appears to be the first author to consider outliers within time series, proposing two types of classes: the additive outliers, which affect only a

single observation, and innovation outliers, which affect succeeding observations. However, the additive outliers deserve special attention, as they usually cause more prejudice in practical problems in time series either with the discreet or continuous marginal distribution in different scenarios. For example, Muler and Yohai (2008) demonstrated that additive outliers could significantly influence GARCH models. They introduced the class of BM-estimates, which limits the propagation of the outlier effect. Similar findings were reported by Elsaied and Fried (2021), who used an iterative approach to derive robust estimators for the INARCH model, and by Kitromilidou and Fokianos (2016), who investigated robust versions of the maximum likelihood estimator with different interventions. Li et al. (2016) robustified the closed-form moment estimators in INGARCH(1,1) models. Kim and Lee (2017) developed robust versions of closed-form moment estimators for INGARCH(1,1) models. Bourguignon and Vasconcellos (2018) assessed the impact of additive outliers in the INAR(1) process. Barczy et al. (2009) and Barczy et al. (2010) conducted theoretical studies on integer-valued autoregressive models of order one contaminated with innovations and additive outliers. In the standard linear time series, such as the ARMA model, Reisen et al. (2020) present an overview of the robust methods and Solci et al. (2023) consider the M-periodogram estimator to obtain confidence intervals using the local bootstrap.

The non-robustness of the maximum likelihood estimator in generalized linear models has been extensively studied in the literature, see Carroll and Welsh (1986), Künsch et al. (1989), and Ruckstuhl and Welsh (1999), among others. Hence, robust estimation procedures have been developed, e.g. Cantoni and Ronchetti (2001), Lo and Ronchetti (2009), and Valdora and Yohai (2014). The work of Cantoni and Ronchetti (2001) is probably one of the most cited, which is based on the quasi-likelihood functions. The authors proposed the Mallows quasi-likelihood estimator (MQLE) considering the class of M-estimators of Mallows (see Mallows (1975)). Outlying observations are bounded separately by weight functions on covariates and by a loss function on the response variable in this method. Although proposed for independent observations, Kitromilidou and Fokianos (2016) extended this method to count time series in the context of the log-linear Poisson model. They found that the MQLE behaved comparably to the classic log-linear model without perturbations. Procedures derived from M-estimators (Huber (1964)) are appropriate alternatives to modeling time series contaminated by outliers or generated by probability distribution with heavy tails; see Bai et al. (1992), Li (2008), and Wu (2007).

Thus, considering the previous discussion, the GLARMA model structure, and the nature of the data application, this paper proposes a robust alternative for the GLARMA Poisson model based on the MQLE estimator. To our best knowledge, robustified proposals for the GLARMA model using M-estimators has never been yet suggested in the literature. Thus, this paper aims to fill this gap. Due to the limitations regarding the asymptotic properties of the GLARMA model, we considered the development of an asymptotic theory for the proposed robust approach beyond the scope of this work. In fact, Davis et al. (2021) claim that theoretical properties for the classic GLARMA model were only established for very restrictive special cases after all these years. However, although a general asymptotic theory has not yet been developed, the finite sample size investigated here corroborates that the estimators are asymptotic consistent. In addition, a Monte Carlo study is realized to evaluate the impact of additive outliers in the response variable and covariates on the classic estimation of a GLARMA model proposed by Davis et al. (2003) and on the robust proposal under distinct scenarios and sample sizes. To show the usefulness of the methodology in real data, we consider the study of the effect of Particulate Material (PM<sub>10</sub>) on the deaths caused by respiratory diseases in Vitória, Brazil.

This work is organized as follows. Section 2 introduces the GLARMA model. Section 3 proposes a robust approach for the GLARMA Poisson model. Section 4 shows a Monte Carlo empirical study to evaluate the performance of the proposed procedure. Section 5 is devoted to a real data analysis, which is the primary motivation of this paper. Finally, Section 6 concludes this work.

## 2 The GLARMA model

The GLARMA models are a class of observation-driven non-Gaussian state space models in which the state process is linearly correlated to the explanatory variables and is non-linearly related to the past values of the observed process (Davis et al., 2003).

Let  $\{Y_t\}$ ,  $t \in \mathbb{Z}$ , be the response series,  $\{X_{i,t}\}$ ,  $1 \leq i \leq k$ , be the  $k$  covariates, and  $\mathcal{F}_{t-1} = \sigma\{Y_s, s \leq t-1; X_{i,s}, 1 \leq i \leq k, s \leq t\}$  be the past of the observed count time process and the past and present of the regressor variables. We assume that the conditional distribution of  $Y_t|\mathcal{F}_{t-1}$  belongs to the exponential family, with density

$$f(y_t|w_t) = \exp\{y_t w_t - b(w_t) + c(y_t)\}, \quad (1)$$

where  $b(\cdot)$  and  $c(\cdot)$  are known real functions, with  $c(\cdot)$  frequently depending on  $y_t$ , and  $\{W_t\}$  summarizes the information in  $\mathcal{F}_{t-1}$ , see (Davis and Dunsmuir, 2016; Davis et al., 2021). Based on (1), the mean and variance of  $Y_t|\mathcal{F}_{t-1}$  are, respectively, given by

$$\mu_t = \mathbb{E}(Y_t|\mathcal{F}_{t-1}) = b'(W_t) \quad \text{and} \quad \sigma_t^2 = \mathbb{V}(Y_t|\mathcal{F}_{t-1}) = b''(W_t),$$

where  $b'$  and  $b''$  refer to the first and second derivative of  $b$ , respectively. The canonical link function  $g$  maps  $\mu_t$  into  $W_t$ , i.e.,  $W_t = g(\mu_t)$ . Therefore,  $g = (b')^{(-1)}$ . The specification of  $W_t$  is given by

$$W_t = \mathbf{X}_t^T \boldsymbol{\beta} + Z_t, \quad (2)$$

where  $\mathbf{X}_t = (1, X_{1,t}, \dots, X_{k,t})^T$ ,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T$  are the regression coefficients, and the noise process  $\{Z_t\}$ , which induces a serial dependence on the observations, is given by

$$Z_t = \sum_{i=1}^{\infty} \gamma_i e_{t-i}, \quad (3)$$

where  $\{e_t\}$  is a martingale difference sequence. The parameters  $\gamma_i$ 's are the coefficients in the following power series expansion satisfying  $\sum_{i=1}^{\infty} |\gamma_i| < \infty$

$$\sum_{i=1}^{\infty} \gamma_i z^i = \frac{\theta(z)}{\phi(z)} - 1, \quad |z| \leq 1, \quad (4)$$

where the autoregressive and moving average polynomials  $\phi(z) = (1 - \phi_1 z - \dots - \phi_p z^p)$  and  $\theta(z) = (1 + \theta_1 z + \dots + \theta_q z^q)$  have no common zeroes and have all their zeros outside the closed unit disk. The model specified by (1), (2), (3) and (4) is called a GLARMA( $p, q$ ) model. From (3) and (4),  $Z_t$  can be calculated recursively with the difference equation

$$Z_t = \phi_1(Z_{t-1} + e_{t-1}) + \dots + \phi_p(Z_{t-p} + e_{t-p}) + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q}. \quad (5)$$

The residuals  $\{e_t\}$  in (3) are given by

$$e_t = \frac{Y_t - \mu_t}{\sigma_t}, \quad (6)$$

It is easy to show that for the Poisson case  $\mathbb{E}(e_t|\mathcal{F}_{t-1}) = 0$  and  $\mathbb{E}(e_t^2|\mathcal{F}_{t-1}) = 1$ . For more details, see Davis et al. (2003).

Define  $\boldsymbol{\delta} = (\beta_0, \beta_1, \dots, \beta_k, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T$  as the parameter vector of model (2). The conditional log-likelihood of  $\{Y_1, \dots, Y_n\}$  is given by

$$L(\boldsymbol{\delta}) = \sum_{t=1}^n \{Y_t W_t(\boldsymbol{\delta}) - b(W_t(\boldsymbol{\delta})) + c(Y_t)\},$$

where

$$W_t(\boldsymbol{\delta}) = \mathbf{X}_t^T \boldsymbol{\beta} + \sum_{i=1}^{\infty} \gamma_i e_{t-i}(\boldsymbol{\delta}), \quad e_t(\boldsymbol{\delta}) = (Y_t - \mu_t(\boldsymbol{\delta})) / \sigma_t(\boldsymbol{\delta}).$$

In the following, we assume that  $Y_t | \mathcal{F}_{t-1}$  follows a Poisson distribution. Then  $W_t = \ln(\mu_t)$ ,  $\sigma_t^2 = \mu_t$ ,  $L(\boldsymbol{\delta})$  is given by

$$L(\boldsymbol{\delta}) = \sum_{t=1}^n \{Y_t W_t(\boldsymbol{\delta}) - \exp(W_t(\boldsymbol{\delta})) - \ln(Y_t!)\}, \quad (7)$$

and can be maximized using Newton-Raphson iterations or Fisher scoring procedure from suitable initial values by computing the first and second derivatives. The conditional maximum likelihood estimator (CMLE) of  $\boldsymbol{\delta}$  maximizes (7). See Davis et al. (2005) for more details about maximization of equation (7).

### 3 Robust estimation for GLARMA models

To robustify the parameter estimation of the GLARMA model, we propose here an extension of MQLE of Cantoni and Ronchetti (2001). Their approach is based on natural generalizations of quasi-likelihood functions, considering a general class of M-estimators of Mallows's type (Mallows, 1975), where the influence of deviations on response variable and covariates are bounded separately.

The MQLE of  $\boldsymbol{\delta}$  is defined as a solution of the estimating equations

$$\sum_{t=1}^n \boldsymbol{\psi}(Y_t, \mu_t(\boldsymbol{\delta})) = \mathbf{0}, \quad (8)$$

where

$$\begin{aligned} \boldsymbol{\psi}(Y_t, \mu_t) &= \nu(Y_t, \mu_t) w(\mathbf{X}_t) \frac{\partial \mu_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} - a(\boldsymbol{\delta}), \\ a(\boldsymbol{\delta}) &= \frac{1}{n} \sum_{t=1}^n \mathbb{E}(\nu(Y_t, \mu_t) | \mathcal{F}_{t-1}) w(\mathbf{X}_t) \frac{\partial \mu_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}. \end{aligned}$$

This is an M-estimator with score function  $\boldsymbol{\psi}(\cdot, \cdot)$  (Huber, 1981; Hampel et al., 1986),  $a(\boldsymbol{\delta})$  is a bias correction used to ensure Fisher's consistency. Function  $\nu(\cdot, \cdot)$  is chosen to control deviations in the  $Y$ -space, and leverage points on  $\mathbf{X}_t$ -space are down-weighted by  $w(\cdot)$ . Both functions  $\nu(\cdot, \cdot)$  and  $w(\cdot)$  are bounded to restrict outlying values on the response variable and covariates, respectively.

We consider the case where  $\nu(Y_t, \mu_t) = \psi_H(e_t) / \sigma_t = \psi_H(e_t) / \mu_t^{1/2}$  with  $\psi_H$  the Huber loss function (Maronna et al. (2006), pages 26-27) defined by

$$\psi_H(x) = \begin{cases} x & \text{if } |x| \leq c, \\ c \operatorname{sign}(x) & \text{if } |x| > c, \end{cases} \quad (9)$$

The Huber function is chosen here because it satisfies assumptions (A1)-(A4) of Reisen et al. (2020), i.e.  $\psi_H(x)$  is a bounded non-decreasing real-valued function on  $\mathbb{R}$ , and  $\lambda(x) = \mathbb{E}[\psi_H(x)(e_1 - x)]$  is well defined, strictly decreasing on  $\mathbb{R}$  and  $\lambda(0) = 0$ . For a more precise definition see, e.g. Reisen et al. (2020); Maronna et al. (2006), among others.

The constant  $c$  must be prespecified and regulates the trade-off between the efficiency and robustness of the estimator. Appropriate choices for  $c$  are in the interval  $[1, 2]$ , and according to Huber (1964),  $c = 1.345$  provides 90% efficiency when the data is normally distributed see, for example, pages 26 and 27, in Maronna et al. (2006). Other specific values are also used in the literature, e.g.

$c = 1.2$  Cantoni and Ronchetti (2001),  $c = 1.25$  Streett et al. (1988); Chi (1994). The choice of  $c$  should reflect the proportion of outliers in the data and must be adjusted according to the data distribution. In the recent paper by Lévy-Leduc et al. (2022), the authors also explored different values of  $c$  in stationary time series to empirically verify the effect of  $c$  on the estimation of the autocovariance function and the spectral density of the process. As expected, for large  $c$ , the Huber function behaves similarly to the standard non-robust estimation methods. In this context, the Huber function becomes useless.

We deduce from (9) that

$$\begin{aligned}\mathbb{E}(\psi_H(e_t)|\mathcal{F}_{t-1}) &= c[\mathbb{P}(Y_t \geq j_{2,t} + 1|\mathcal{F}_{t-1}) - \mathbb{P}(Y_t \leq j_{1,t}|\mathcal{F}_{t-1})] \\ &\quad + \mu_t^{1/2}[\mathbb{P}(Y_t = j_{1,t}|\mathcal{F}_{t-1}) - \mathbb{P}(Y_t = j_{2,t}|\mathcal{F}_{t-1})],\end{aligned}$$

where  $j_{1,t} = \lfloor \mu_t - c\mu_t^{1/2} \rfloor$  and  $j_{2,t} = \lfloor \mu_t + c\mu_t^{1/2} \rfloor$ . Moreover, since  $\mu_t = \exp(W_t)$ ,  $\partial\mu_t(\boldsymbol{\delta})/\partial\boldsymbol{\delta} = \mu_t\partial W_t(\boldsymbol{\delta})/\partial\boldsymbol{\delta}$ .

A common choice for  $w(\cdot)$  is  $w(\mathbf{X}_t) = \sqrt{1 - h_t}$ , where  $h_t$  is the  $t$ th component of the diagonal of the projection matrix  $H = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}$  and  $\mathbf{X}$  is the  $(n \times (k+1))$  matrix defined by  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_n]^T$ , see (Cantoni and Ronchetti, 2001). However,  $w(\mathbf{X}_t)$  does not have high breakdown properties and more sophisticated weights can be found in the literature.

Let  $\mathbf{V}_t = (X_{1,t}, \dots, X_{k,t})^T$  for  $t = 1, \dots, n$ ,  $\bar{\mathbf{V}}$  the arithmetic mean of the vectors  $\mathbf{V}_t$  and  $C$  their sample covariance matrix,

$$\bar{\mathbf{V}} = \frac{1}{n} \sum_{t=1}^n \mathbf{V}_t, \quad C = \frac{1}{n-1} \sum_{t=1}^n (\mathbf{V}_t - \bar{\mathbf{V}})^T (\mathbf{V}_t - \bar{\mathbf{V}}).$$

We assume that  $C$  is invertible and we measure how far  $\mathbf{V}_t$  is from  $\bar{\mathbf{V}}$  in the metric defined by  $C$ , yielding

$$\text{MD}_t^2 = (\mathbf{V}_t - \bar{\mathbf{V}})C^{-1}(\mathbf{V}_t - \bar{\mathbf{V}})^T,$$

which is called the squared Mahalanobis distance of  $\mathbf{V}_t$  from  $\bar{\mathbf{V}}$ . To robustify  $\text{MD}_t^2$ ,  $\bar{\mathbf{V}}$  and  $C$  can be estimated using the minimum covariance determinant (MCD) algorithm, see (Rousseeuw and Leroy, 1987, Chapter 7.1). In this procedure,  $h$  observations out of  $n$  are chosen whose classical sample covariance matrix presents the lowest determinant. Then the MCD estimate of location  $\bar{\mathbf{V}}_{\text{MCD}}$  is the arithmetic mean of these  $h$  vectors  $\mathbf{V}_t$  and  $C_{\text{MCD}}$  is their sample covariance matrix. In this paper, we use the weight function  $w(\cdot)$  based on the MCD estimates given by

$$w(\mathbf{X}_t) = \min \left[ 1, \left\{ \frac{b}{(\mathbf{V}_t - \bar{\mathbf{V}}_{\text{MCD}})^T C_{\text{MCD}}^{-1} (\mathbf{V}_t - \bar{\mathbf{V}}_{\text{MCD}})} \right\}^{\alpha/2} \right], \quad (10)$$

where  $\alpha$  and  $b$  are tuning constants. Simpson et al. (1992) evaluate some values for the constant  $\alpha$  and claim that  $\alpha = 1$  is usual for the class of generalized M-estimators. In addition, the authors set  $b$  equal to the  $(1 - \gamma)$ -quantile of the chi-squared distribution with  $k - 1$  degrees of freedom, where  $k$  is the number of predictor covariates and  $\gamma = 0.1$  or  $\gamma = 0.05$ .

## 4 Simulation study

We compare the performances of the CMLE and the MQLE of  $\boldsymbol{\delta} = (\beta_0, \beta_1, \dots, \beta_k, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T$  using the GLARMA(1,0) Poisson model defined by

$$\begin{aligned}Y_t|\mathcal{F}_{t-1} &\sim \text{Poisson}(\mu_t), \\ \ln(\mu_t) &= \beta_0 + \beta_1 X_t + Z_t, \\ X_t &= \xi X_{t-1} + U_t,\end{aligned} \quad (11)$$

where  $\{U_t\}$  is a sequence of independent zero-mean, unit-variance Gaussian random variables. We set  $\beta_0 = 1$ ,  $\beta_1 = 0.5$ , and  $X_t$  is an autoregressive process of order 1. We consider two scenarios for the process  $X_t$ ;  $\xi = 0$  (Gaussian white noise process) and  $\xi = 0.4$ .

The Monte Carlo simulations are repeated 1000 times with sample sizes equal to  $n = 100$  and  $n = 1000$  and the results in Tables 2 to 8 refer to the sample Mean and Mean Square Error (MSE).

This section is organized as follows: Subsection 4.1 covers the selection of  $c$  value, while subsection 4.2 describes the situation in which the covariate  $X_t$  is affected by additive outliers. Subsection 4.3 evaluates the scenario in which outliers contaminate the response variable  $Y_t$ . Finally, subsection 4.4 discusses the scenario in which both the covariate and the response variable are contaminated.

## 4.1 The choice of $c$

Regarding selecting the constant  $c$ , we investigated two approaches. First, we conducted a simulation study to determine the values of  $c$  that resulted in the smallest measures of MSE. We performed 1000 simulations and identified that values of the tuning parameter between 1.0 and 4.0 were the most common. The GLARMA(1,0) model defined by equation (11) was fitted with  $\xi = 0$ , and  $n = 100$ . We opted for this simplest case, with a small sample size, due to the computational cost associated with this procedure and to verify the empirical properties even for small sample sizes. Table 1 presents the ten most common values of  $c$ . The best adjustments were observed for  $c$  values of 1.6, 4.0, and 3.5. It is important to note that the percentages of these values are not very high. The most common value in the simulation study,  $c = 1.6$ , represents only approximately 5% of the cases that led to the best adjustment. Additionally, the  $c$  values in Table 1 appear somewhat random, as no clear trend is evident among them.

Table 1: Top 10 most common values of  $c$  in the simulation study - GLARMA(1,0) model with  $\xi = 0$ ,  $\beta_0 = 1.0$ ,  $\beta_1 = 0.5$  -  $n = 100$

The percentage represents the proportion in which each value of  $c$  leads to the best adjustments.

$c$	1.6	4.0	3.5	3.4	3.3	2.7	2.4	1.4	3.6	2.5	3.2
%	4.8	4.4	4.2	4.1	4.1	4.0	3.8	3.8	3.6	3.6	3.4

In the second approach, we employed a cross-validation procedure for time series using blocks, as described in Bergmeir and Benitez (2012), and Bergmeir et al. (2018). Cross-validation is a resampling procedure used to evaluate machine-learning methods. In this case, the algorithm selects the value of  $c$  that yields the smallest values of the MSE.

Table 2 summarizes the parameter estimation and the MSE for the CMLE GLARMA(1,0) model, as well as the robust proposal (MQLE) using the approaches discussed in this subsection: cross-validation with two blocks and a fixed constant  $c$  at values 1.6, 4.0, and 3.5 (the first three most frequent tuning parameter values from Table 1). Additionally, we include the estimations for  $c = 1.345$ , which, according to Huber (1981), provides 90% efficiency for normal data. We found that in cross-validation, the MSE is similar to that of the classic method when no outliers are present. However, with outliers, the mean of parameter  $\beta_1$  is underestimated, as well as in the classic method. When examining distinct values of  $c$ , we observed that the MSE is quite similar for clean data. For perturbed data, the mean of  $\beta_1$  was underestimated for  $c = 3.5$  and  $c = 4.0$ . The MSE and the mean of  $\beta_1$  were quite similar for  $c = 1.345$  and  $c = 1.6$ . Since these values are very close, we have decided to set  $c = 1.345$  in equation (9), as this value has been consistently cited in the literature and provided the smallest measures of MSE for  $\beta_1$  in clean and perturbed data (Table 2). Note that, for  $\beta_1$ , the MSE of the CMLE estimates increase by almost 14% when outliers are present, compared to the estimations without contamination. In contrast, the MSE of the robust proposal with  $c = 1.345$  changes by less than 1%.



Table 2: Parameters estimation GLARMA(1,0),  $\beta_0 = 1.0$ ,  $\beta_1 = 0.5$ ,  $\xi = 0$ ,  $\omega = 5$ ,  $\phi = 0.2$  - n=100

Methods for choosing $c$		no outlier		with outlier	
		Mean	MSE	Mean	MSE
CMLE	$\beta_0$	0.9741	0.0070	1.0333	0.0065
	$\beta_1$	0.4980	<b>0.0035</b>	0.2797	<b>0.0494</b>
	$\phi$	0.1303	0.0085	0.0834	0.0184
Crossvalidation 2 blocks	$\beta_0$	0.9695	0.0076	1.0004	0.0060
	$\beta_1$	0.5067	0.0039	0.4154	0.0103
	$\phi$	0.1484	0.0075	0.1338	0.0114
$c = 1.345$	$\beta_0$	0.9635	0.0076	0.9847	0.0065
	$\beta_1$	0.5259	<b>0.0047</b>	0.4674	<b>0.0041</b>
	$\phi$	0.1827	0.0064	0.1779	0.0069
$c = 1.6$	$\beta_0$	0.9639	0.0082	0.9854	0.0065
	$\beta_1$	0.5240	0.0049	0.4647	0.0045
	$\phi$	0.1778	0.0063	0.1728	0.0069
$c = 4.0$	$\beta_0$	0.9655	0.0081	0.9878	0.0064
	$\beta_1$	0.5173	0.0044	0.4546	0.0049
	$\phi$	0.1636	0.0065	0.1574	0.0073
$c = 3.5$	$\beta_0$	0.9740	0.0083	1.0107	0.0058
	$\beta_1$	0.4980	0.0046	0.3870	0.0148
	$\phi$	0.1302	0.0085	0.1059	0.0129

## 4.2 Scenario 1: Covariate contaminated by additive outliers

We compare the performances of the CMLE and the MQLE of  $\delta$  using the GLARMA(1,0) Poisson model in equation (11).

The contaminated version of  $X_t$  is defined by  $X_t^* = X_t + \omega\varphi_t$ , where  $\omega = 5$  is the magnitude of the outlier, which impacts  $X_t$  and  $\varphi_t$  indicates the presence or not of this outlier and its sign at the time  $t$ , i.e.  $\varphi_t = 0$  with probability  $1 - \varphi$ , and  $\varphi_t = \pm 1$  with probability  $\varphi/2$  where  $\varphi = 0.01$ .

Here, we focus on a GLARMA(1,0) model, as in practice, most data can be adequately modeled with just an AR(1) process as the autocorrelation structure. Davis et al. (2003) developed appropriate asymptotic theory for GLARMA(0,1). The empirical study involving this model is discussed in the Appendix.

### 4.2.1 GLARMA(1,0) model

The GLARMA(1,0) model is defined by (11), where  $\{Z_t\}$  satisfies

$$Z_t = \phi[Z_{t-1} + (Y_{t-1} - \exp(W_{t-1})) \exp(-W_{t-1}/2)]. \quad (12)$$

We set  $\phi = 0.2$  because previous studies have shown that increasing the value of  $\phi$  can impact the parameter estimation in GLARMA models. In a still unpublished paper, the authors investigate various values of the parameter  $\phi$  to understand its impact in this family model.

Table 3 presents the parameter estimation results when the covariate  $X_t$  is an independent random vector ( $\xi = 0$ ). For both sample sizes, in the absence of outliers, the CMLE of  $\beta_0$  and  $\beta_1$  are close to the true values, while  $\phi$  is underestimated. In the presence of additive outliers, the CMLE of  $\beta_1$  and  $\phi$  are underestimated and have larger MSE. Without contamination, the MQLE is close to the CMLE, except for the estimation of  $\phi$ , which is closer to the true value and displays a smaller MSE. In the contaminated case, the MQLE are close to the true values, and their MSE are comparable to the ones of the CMLE in the absence of outliers.

Table 3: GLARMA(1,0) model with  $\xi = 0$ ,  $\beta_0 = 1.0$ ,  $\beta_1 = 0.5$ ,  $\phi = 0.2$ , and  $\omega = 5.0$

Estimation methods		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.973	0.0071	1.029	0.0073	0.995	0.0006	1.020	0.0009
	$\hat{\beta}_1$	0.501	0.0029	0.404	0.0139	0.505	0.0003	0.421	0.0064
	$\hat{\phi}$	0.119	0.0100	0.082	0.0166	0.127	0.0057	0.115	0.0077
MQLE	$\hat{\beta}_0$	0.965	0.0082	0.972	0.0078	0.983	0.0009	0.992	0.0007
	$\hat{\beta}_1$	0.536	0.0049	0.507	0.0041	0.530	0.0012	0.502	0.0003
	$\hat{\phi}$	0.169	0.0064	0.160	0.0069	0.178	0.0009	0.176	0.0011

Table 4 presents the parameter estimation results when  $X_t$  is an AR(1) process ( $\xi = 0.4$ ). As observed in Table 3, the CMLE with clean data are close to the true values, except for  $\phi$ , which is underestimated. In the presence of additive outliers, the CMLE of  $\beta_1$  and  $\phi$  are affected. Observe that for  $n = 100$ , the estimate of  $\beta_1$  is 0.299 (Table 4) in comparison with 0.404 when  $\xi = 0$  (Table 3). Then the time correlation of the affected covariate seems to be confounding. In the absence of outliers, the MQLE are close to the true values with MSE slightly larger than the CMLE, except for the estimate of  $\phi$ . In the contaminated case, the MQLE are close to the true values and have similar MSE as the CMLE in the absence of outliers. The same behaviors are observed for  $n = 100$  and  $n = 1000$ . As the number of observations  $n$  increases, the MSE decreases.

Table 4: GLARMA(1,0) model with  $\xi = 0.4$ ,  $\beta_0 = 1.0$ ,  $\beta_1 = 0.5$ ,  $\phi = 0.2$ , and  $\omega = 5.0$

Estimation methods		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.981	0.0069	1.005	0.0059	0.997	0.0006	1.041	0.0022
	$\hat{\beta}_1$	0.511	0.0042	0.299	0.0414	0.503	0.0003	0.376	0.0153
	$\hat{\phi}$	0.117	0.0099	0.135	0.0081	0.119	0.0068	0.108	0.0088
MQLE	$\hat{\beta}_0$	0.967	0.0079	0.983	0.0068	0.982	0.0009	0.993	0.0006
	$\hat{\beta}_1$	0.549	0.0072	0.482	0.0038	0.528	0.0012	0.499	0.0003
	$\hat{\phi}$	0.169	0.0059	0.168	0.0066	0.167	0.0015	0.166	0.0016

### 4.3 Scenario 2: Response variable contaminated by additive outliers

Now, the response variable  $\{Y_t\}$  is impacted by additive outliers. The contaminated version of  $Y_t$  is defined by  $Y_t^* = Y_t + \omega\varphi_t$ , where  $\omega = 30$  is the magnitude of the outlier which impacts  $Y_t$  and  $\varphi_t$  indicates the presence or not of this outlier at time  $t$ , i.e.  $\varphi_t = 0$  with probability  $1 - \varphi$ , and  $\varphi_t = 1$  with probability  $\varphi$  where  $\varphi = 0.01$ . Note that the magnitude of  $\omega$  varies across scenarios. In the Gaussian process,  $\omega = 5$  is large enough to contaminate the covariate. However, the response variable follows the Poisson distribution, where the mean equals the variance. In this case, the variability is much larger than the variance of the process  $X_t$ , and the magnitude of 5 becomes too small to contaminate  $Y_t$ .

The simulation study related to the GLARMA(0,1) model is in the Appendix.

#### 4.3.1 GLARMA(1,0) model

Now, we take model (11) where  $\{Z_t\}$  is defined by (12) with  $\phi = 0.2$ .

Table 5 presents the parameter estimation results when  $\xi = 0$ . Without outlier, the CMLE of  $\beta_0$  and  $\beta_1$  are close to the true values, and  $\theta$  is underestimated for both sample sizes. In the presence of additive outliers on  $\{Y_t\}$ , all the CMLE are impacted for  $n = 100$  and  $n = 1000$ . Without outlier, the MQLE is globally less efficient than the CMLE. In the contaminated case, the MQLE are globally unaffected and close to the true values. These results apply for  $n = 100$  and  $n = 1000$ .

Table 5: GLARMA(1,0) model with  $\xi = 0$ ,  $\beta_0 = 1.0$ ,  $\beta_1 = 0.5$ ,  $\phi = 0.2$ , and  $\omega = 30.0$

Estimation methods		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.975	0.0068	1.229	0.0565	0.996	0.0006	1.183	0.0340
	$\hat{\beta}_1$	0.506	0.0024	0.423	0.0076	0.503	0.0002	0.431	0.0049
	$\hat{\phi}$	0.121	0.0093	0.111	0.0080	0.127	0.0056	0.010	0.0359
MQLE	$\hat{\beta}_0$	0.963	0.0081	1.004	0.0065	0.985	0.0008	1.000	0.0005
	$\hat{\beta}_1$	0.527	0.0036	0.509	0.0028	0.528	0.0011	0.525	0.0009
	$\hat{\phi}$	0.167	0.0059	0.170	0.0050	0.176	0.0010	0.168	0.0014

Table 6 presents the parameter estimation results when  $\xi = 0.4$ . As observed in Table 5, without perturbation, the CMLE is close to the true values, except for  $\phi$  which is underestimated. In the presence of outliers, all CMLE are affected. In the absence of outliers, the MQLE are close to the true values, with the MSE slightly larger than the MSE of the CMLE, except for  $\phi$ . In the presence of outliers, the MQLE are close to the true values and have similar MSE as the CMLE in the absence of outliers. Similar results are observed for  $n = 100$  and  $n = 1000$ .

Table 6: GLARMA(1,0) model with  $\xi = 0.4$ ,  $\beta_0 = 1.0$ ,  $\beta_1 = 0.5$ ,  $\phi = 0.2$ , and  $\omega = 30.0$

Estimation methods		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.975	0.0067	1.28	0.0824	0.997	0.0006	1.118	0.0146
	$\hat{\beta}_1$	0.501	0.0032	0.413	0.0097	0.501	0.0003	0.448	0.0029
	$\hat{\phi}$	0.116	0.0105	-0.007	0.0439	0.117	0.0071	0.021	0.0319
MQLE	$\hat{\beta}_0$	0.966	0.0077	0.988	0.0063	0.982	0.0010	0.993	0.0007
	$\hat{\beta}_1$	0.525	0.0043	0.530	0.0048	0.525	0.0009	0.523	0.0009
	$\hat{\phi}$	0.160	0.0069	0.166	0.0076	0.165	0.0016	0.162	0.0018

#### 4.4 Scenario 3: Covariate and Response variable contaminated by additive outliers

In this subsection, both the covariate and the response variable  $Y_t$  are affected by additive outliers. As described in section 4.2, the contaminated version of  $X_t$  is defined by  $X_t^* = X_t + \omega_1 \varphi_t$ , where  $\omega_1 = 5$  represents the magnitude of the outlier impacting  $X_t$ , and  $\varphi_t$  indicates the presence or absence of this outlier and its sign at time  $t$ . Specifically,  $\varphi_t = 0$  with probability  $1 - \varphi$ , and  $\varphi_t = \pm 1$  with probability  $\varphi/2$ , where  $\varphi = 0.01$ . The contaminated version of  $Y_t$  is defined by  $Y_t^* = Y_t + \omega_2 \varphi_t$ , where  $\omega_2 = 30$  represents the magnitude of the outlier impacting  $Y_t$ , and  $\varphi_t$  indicates the presence or absence of this outlier at time  $t$ . In this context, we consider the GLARMA(1,0) model defined by (11) where  $\{Z_t\}$  is defined by (12) with  $\phi = 0.2$ .

Table 7 presents the parameter estimation results when  $\xi = 0$ . For  $n = 100$  and no outlier, the CMLE of  $\beta_0$  and  $\beta_1$  are close to the true values, and  $\phi$  is underestimated. With additive outliers on  $(Y_t)$ , all CMLE are impacted:  $\beta_0$  is overestimated,  $\beta_1$  and  $\phi$  are underestimated, and all MSE increase. Without additive outliers on  $\{Y_t\}$ , the MQLE of  $\beta_0$  and  $\beta_1$  are close to the CMLE, and the estimate of  $\phi$  is closer to the true value with a smaller MSE. For  $n = 100$  and additive outliers, contrarily to the CMLE, and the MQLE are unaffected and provide estimates close to the true parameters. Similar conclusions are observed for  $n = 1000$ . As the  $n$  increases, the MSE decreases.

Table 7: GLARMA(0,1) model with  $\xi = 0$ ,  $\beta_0 = 1.0$ ,  $\beta_1 = 0.5$ ,  $\phi = 0.2$ ,  $\omega_1 = 5.0$ ,  $\omega_2 = 30.0$

Estimation methods		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.977	0.0071	1.121	0.0194	0.996	0.0006	1.137	0.0193
	$\hat{\beta}_1$	0.526	0.0047	0.388	0.0153	0.504	0.0003	0.468	0.0012
	$\hat{\phi}$	0.120	0.0099	0.009	0.0371	0.125	0.0058	0.031	0.028
MQLE	$\hat{\beta}_0$	0.968	0.0077	0.977	0.0071	0.982	0.0009	0.997	0.0006
	$\hat{\beta}_1$	0.539	0.0053	0.526	0.0047	0.533	0.0014	0.531	0.0013
	$\hat{\phi}$	0.171	0.0073	0.168	0.0060	0.172	0.0012	0.165	0.0017

Table 8 presents the parameter estimation results when  $\xi = 0.4$ . Similar to the previous scenarios, without perturbation, for  $n = 100$  the CMLEs are close to the true values, except for  $\phi$ , which is underestimated. For  $n = 1000$  the CMLE of  $\beta_0$  and  $\beta_1$  are very close to the true values but  $\phi$  is still underestimated. When additive outliers are present, the CMLE parameters are affected for  $n = 100$  and  $n = 1000$ . In the absence of outliers, the MQLEs provide good results, with the MSE slightly larger than that of the CMLEs, except for  $\phi$ . In the presence of outliers, the MQLEs give almost the same results as the CMLE in the absence of outliers.

Table 8: GLARMA(1,0) model with  $\xi = 0.4$ ,  $\beta_0 = 1.0$ ,  $\beta_1 = 0.5$ ,  $\phi = 0.2$ ,  $\omega_1 = 5.0$ , and  $\omega_2 = 30.0$

Estimation methods		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.979	0.0062	1.081	0.011	0.999	0.0007	1.15	0.0231
	$\hat{\beta}_1$	0.499	0.0046	0.155	0.121	0.501	0.0003	0.363	0.0189
	$\hat{\phi}$	0.13	0.0085	0.017	0.035	0.115	0.0074	0.022	0.0319
MQLE	$\hat{\beta}_0$	0.973	0.0067	0.974	0.0067	0.983	0.0009	1.003	0.0007
	$\hat{\beta}_1$	0.471	0.0058	0.471	0.0059	0.528	0.0011	0.504	0.0004
	$\hat{\phi}$	0.183	0.0069	0.203	0.0039	0.162	0.0018	0.154	0.002

This empirical study shows that the CMLE are impacted by additive outliers, both in the covariates and in the response variable. The MQLE are similar to the CMLE in the absence of additive outliers. With contaminated data, the MQLE outperform the CMLE, providing parameter estimates closer to the true values with smaller MSE. In all scenarios analyzed, the MSE decreases as the sample size increases.

## 5 Real data analysis

We evaluate the impact of particulate matter (PM<sub>10</sub>) on the monthly number of deaths by respiratory diseases between 2011 to 2018 ( $n = 96$ ) in Greater Vitória, Brazil, which is a port and

industrialized region, densely populated in the state of Espírito Santo, with approximately 1,900,000 inhabitants. Although the atmosphere is composed of many gases and particulate matter, only PM<sub>10</sub> is considered because the data quality of other contaminants during the period of study is too poor. PM<sub>10</sub> are microscopic solid particles and liquid droplets suspended in the air, with a diameter of 10 micrometers or less. This particle pollution mainly comes from motor vehicles, wood-burning heaters, and industry. It has been associated with premature mortality, increased hospital admissions for heart or lung causes, acute and chronic bronchitis, asthma attacks, and respiratory symptoms (Schwartz, 2000).

The correlation between the number of deaths and the maximum monthly concentrations of PM<sub>10</sub> in the atmosphere is  $\rho = 0.45$ , which is significant. Before fitting the model, some data were imputed to handle the missing observations in the PM<sub>10</sub> series. We have used the multivariate imputation by chained equation method proposed by van Buuren and Oudshoorn (2000). Using imputation-based procedures can have several implications. van Buuren and Oudshoorn (2011) outlines the primary issues related to imputing missing data in multivariate contexts. In a review of imputation methods for time series in epidemiological studies, Junger and De Leon (2015) emphasizes the necessity of these procedures when faults in air quality measurement devices render them inoperable. Gaps in data can lead to distortions in quantifying exposure. However, a key drawback of imputation-based procedures is that they generally do not account for imputation-related imprecision in the analysis, leading to an underestimation of the variance of the estimators.

Figure 1 presents the deaths caused by respiratory diseases and PM<sub>10</sub> concentrations. The number of deaths shows a positive trend and seasonal behavior. The PM<sub>10</sub> concentration also presents a positive trend and three peaks. These aberrant observations can be considered additive outliers. The peaks in the data are associated with specific characteristics. As noted in Huebnerova and Michalek (2014) among others, dust aerosol levels can be influenced by various factors like the weekday, the heating season, cloud cover, and wind speed. Atypical values may stem from very localized sources, such as a small fire or a truck idling near a monitoring station, or from unusual weather conditions like low mixing height and high atmospheric stability, which can lead to poor dispersion of emitted pollutants (zan Zoest et al. (2018)).

Our GLARMA Poisson model includes a trend in the number of deaths, and sine and cosine functions are incorporated to handle the annual seasonality. The model is written as

$$W_t = \beta_0 + \beta_1 X_t + \beta_2 t + \beta_3 \sin(2\pi t/12) + \beta_4 \cos(2\pi t/12) + Z_t, \quad (13)$$

where  $t$  is the month number,  $X_t$  is the PM<sub>10</sub> concentration, and  $Z_t$  is the time-correlated noise defined by (5). Table 9 presents the value of the Bayesian information criterium (BIC) for each GLARMA( $p, q$ ) Poisson model for  $0 \leq p, q \leq 4$  with  $p \neq q$ . When  $p = q$ , the estimators' identifiability problems and lack of convergence are known (Dunsmuir and Scott, 2015). The minimum value is obtained for  $(p, q) = (1, 0)$  for both estimation methods.

Tables 10 and 11 present the CMLE and MQLE of  $(\beta_0, \dots, \beta_4)$ , respectively, for the GLARMA(1, 0) model. The Wald test was used to evaluate whether individual coefficients are zero. In Table 10, all the estimates are significant at the 5% level of significance, except the coefficient  $\hat{\beta}_1$  related to the PM<sub>10</sub> levels in the atmosphere. In Table 11, all the estimates are significant at the 5% level of significance. Although there is a significant correlation between the monthly number of deaths in the period and PM<sub>10</sub>, the CMLE of  $\beta_1$  is not significant at 5% level of significance. However, the MQLE of  $\beta_1$  is significant, which means that PM<sub>10</sub> contributes significantly to the increase in deaths caused by respiratory diseases. Observe that the value of  $\beta_1$  seems to be underestimated by CMLE ( $\hat{\beta}_1 = 0.0003$ ) since the robust estimate is three times this value ( $\hat{\beta}_1 = 0.0009$ ).

Standardized Pearson residuals (Harvey and Fernandes (1989)) are commonly used to assess the adequacy of a model. An adequate model should have uncorrelated Pearson residuals with a mean close to 0 and a sample variance close to 1 (Weiss (2018)). Figure 2 displays the behavior of the

Figure 1: Monthly number of deaths by respiratory diseases and maximum monthly concentrations of PM<sub>10</sub> in the Greater Vitória, Brazil.

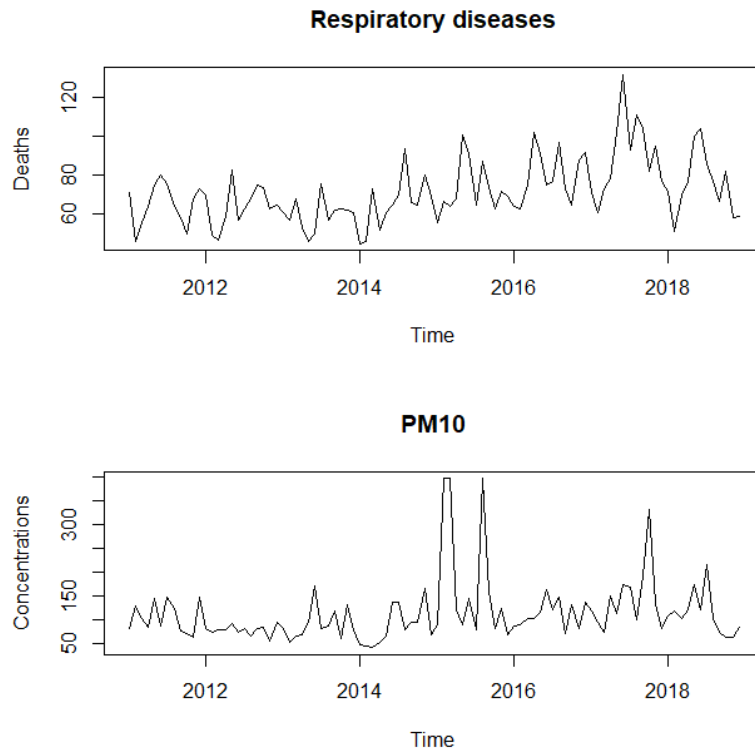


Table 9: BIC values for the GLARMA( $p, q$ ) Poisson model (13).

$(p, q)$	(0, 0)	(0, 1)	(0, 2)	(0, 3)	(0, 4)	(1, 0)	(1, 2)
CMLE	808,132	811,644	812,132	812,503	812,004	<b>793,206</b>	795,716
MQLE	827,829	827,865	827,933	829,685	835,549	<b>806,466</b>	809,496
$(p, q)$	(1, 3)	(1, 4)	(2, 0)	(2, 1)	(2, 3)	(2, 4)	(3, 0)
CMLE	796,709	797,953	812,12	795,50	815,97	815,89	812,50
MQLE	811,376	814,399	827,91	809,67	831,86	834,30	829,63
$(p, q)$	(3, 1)	(3, 2)	(3, 4)	(4, 0)	(4, 1)	(4, 2)	(4, 3)
CMLE	796,18	815,96	816,52	812,00	797,08	815,89	816,52
MQLE	810,12	831,83	837,90	835,51	813,07	834,27	837,98

Table 10: CMLE of model (13).

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Estimate	4.0571	0.0003	0.0034	-0.0573	-0.0954
Standard error	0.0383	0.0002	0.0006	0.0228	0.0228
$p$ value	<2e-16	0.0706	<1.9e-06	0.0121	2e-05

sample ACF and PACF functions of the Pearson residuals of the fitted models, considering CMLE and MQLE estimation methods. Table 12 presents the sample means and variances of the adjusted models. Although the sample variance of the model estimated using CLME is slightly higher than that of the MQLE residual, we can see that both fitted models satisfactorily capture the statistical association between pollutants and health effects, as the residuals are not correlated, i.e. they can

Table 11: MQLE of Model (13).

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Estimate	4.0097	0.0009	0.0030	-0.0495	-0.0824
Standard error	0.0311	0.0001	0.0004	0.0177	0.0180
$p$ value	<2e-16	2e-07	1e-10	0.0053	4e-06

be seen as an approximate white noise process.

Figure 2: Sample ACF and PACF of the Pearson residuals from model (13).

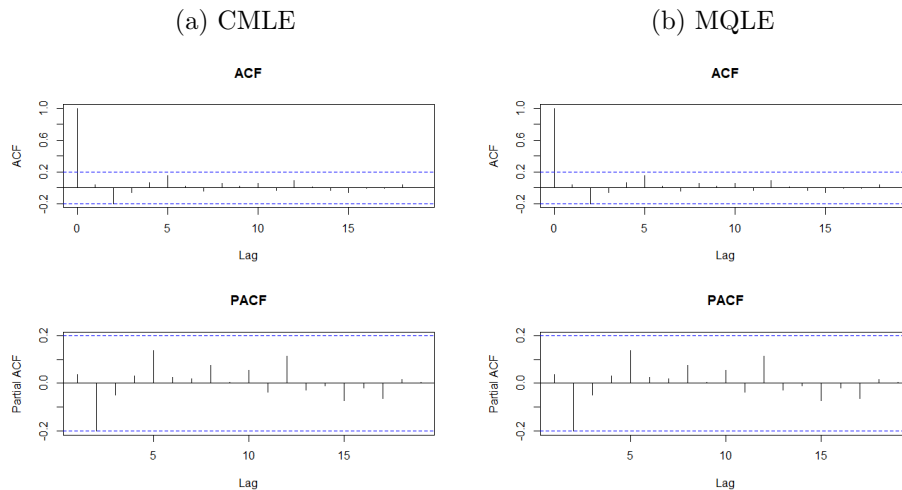


Table 12: Sample mean and variance of the Pearson residuals from model (13)

	CMLE	MQLE
Mean	-0.0099	-0.0693
Variance	1.9219	1.0126

Figure 3 presents the estimated probability integral transform (PIT) of the model (13), based on the conditional cumulative distribution function (Czado et al. (2009)). If the chosen distribution for the model is accurate, the histogram should resemble those obtained when sampling from the uniform distribution on the interval  $[0, 1]$ . Both graphs in Figure 3 suggest deviations from uniformity, indicating that the Poisson model may not adequately capture the data dispersion. Contrary to this finding, the residual analysis indicates that the chosen model is appropriate for estimating the data dynamics. Therefore, the feature in Figure 3 may be attributed to random sampling fluctuations. However, this suggests the need for further investigation. One idea is to compare different loss functions in future works. In addition, we can extend this robust approach to GLARMA models using other distributions, such as the negative binomial, which may better account for data dispersion in this real data set.

As an example in the real scenario to show the effectiveness of the proposed method, the abrupt measurements of the PM10 concentration series were replaced by the mean of the data. As can be seen from Table 13, all CMLE estimates are significant at the 5% level of significance and are very close to the MQLE, i.e. the modified data provides estimates similar to the robust ones of the original data. This simple example demonstrated the importance of using the robust method proposed here when additive outliers are suspected in the data.

Figure 3: Histograms of probability integral transform (PIT) based on Model (13).

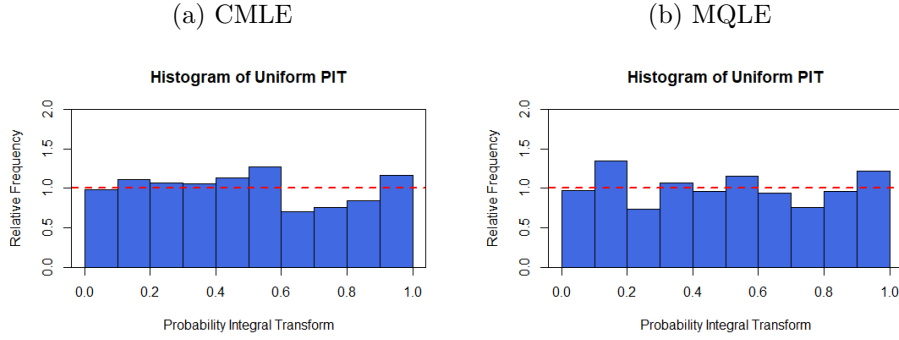


Table 13: CMLE of Model (13) fitted to modified PM<sub>10</sub> concentrations.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Estimate	4.0196	0.0008	0.0033	-0.0540	-0.0874
Standard error	0.0447	0.0003	0.0005	0.0220	0.0225
<i>p</i> value	<2e-16	0.0274	2e-08	0.0143	0.0001

In the epidemiology context, the impact of air pollutants on human health is evaluated by their relative risk (RR). The RR of a variable  $X_i = \{X_{i,t}\}$  is the change in the expected count of the response variable per  $\zeta$ -unit change in  $X_i$ , keeping the other covariates fixed. For Poisson regression, the RR of  $X_i$  is  $RR_{X_i}(\zeta) = \exp(\beta_i \zeta)$ , where  $\beta_i$  is the coefficient associated to the regressor  $X_i$ . An estimate of  $RR_{X_i}(\zeta)$  is

$$\widehat{RR}_{X_i}(\zeta) = \exp(\hat{\beta}_i \zeta), \quad (14)$$

where  $\hat{\beta}_i$  is an estimator of  $\beta_i$ . An estimated asymptotic confidence interval (CI) at an  $\alpha$  significance level of  $RR_{X_i}(\zeta)$  is

$$\widehat{CI}\{RR_{X_i}(\zeta)\} = \left( \exp\{\zeta(\hat{\beta}_i - z_{\alpha/2} \hat{\sigma}(\hat{\beta}_i))\}, \exp\{\zeta(\hat{\beta}_i + z_{\alpha/2} \hat{\sigma}(\hat{\beta}_i))\} \right), \quad (15)$$

where  $\hat{\sigma}(\hat{\beta}_i)$  is the estimated standard deviation of  $\hat{\beta}_i$  and  $z_{\alpha/2}$  denotes the  $(1 - \alpha/2)$ -quantile of the standard normal distribution. Table 14 presents the estimated RR and CI for PM<sub>10</sub> calculated from (14) and (15) with  $\alpha = 5\%$ . The CMLE and MQLE provide significant estimates of the RR since the value one does not belong to the CI.

Table 14: Estimated RR and 95% CI for PM<sub>10</sub> in Model (13).

PM <sub>10</sub>	CMLE	MQLE
$\widehat{RR}$	1.0187	1.0497
$\widehat{CI}$	[1.0001;1.0376]	[1.0305;1.0692]

## 6 Conclusion

This work proposed a robust approach for the GLARMA model. This methodology is based on the robustification of the quasi-likelihood function using M-estimators to control deviations on response variable and weight functions to limit leverage points on covariates.



The simulation showed that additive outliers could widely affect the CMLE. The robust estimates behave approximately like the CMLE in the absence of outliers. For contaminated data, the robust parameter estimation was almost as good as the CMLE in the case of clean observations.

The robust model was applied to the monthly number of deaths caused by respiratory diseases in Vitória, Brazil, to evaluate the impact of PM<sub>10</sub> on the population's health. The observed RR indicated that PM<sub>10</sub> contributed significantly to the increase of deaths by respiratory disease in the region. This analysis also showed that the CMLE underestimated the RR by at least 3%.

## Acknowledgements

The authors thank the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES), National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq), Minas Gerais State Research Foundation (Fundação de Amparo à Pesquisa do Estado de Minas Gerais — FAPEMIG), and Espírito Santo State Research Foundation (Fundação de Amparo à Pesquisa do Espírito Santo — FAPES). This research was also supported by DATAIA convergence institute as part of the “Programme d’Investissement d’Avenir”, (ANR-17-CONV-0003) operated by CentraleSupélec.

The authors thank the referees for the suggestions, which have significantly improved the manuscript.

## References

- Bai, Z., Rao, C., and Wu, Y. (1992). M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Stat. Sin.*, 2(1):237–254.
- Ballesteros-Cánovas, J., Trappmann, D., Madrigal-González, J., Eckert, N., and Stoffel, M. (2018). Climate warming enhances snow avalanche risk in the Western Himalayas. *Proc. Natl. Acad. Sci. USA*, 115(13).
- Barczy, M., Ispány, M., Pap, G., Scotto, M., and Silva, M. (2009). Additive outliers in INAR(1) models. *Stat. Pap.*, 53.
- Barczy, M., Ispány, M., Pap, G., Scotto, M., and Silva, M. (2010). Innovational outliers in INAR(1) models. *Commun. Stat. Theory Methods*, 39:3343–3362.
- Benjamin, M., Rigby, R., and Stasinopoulos, D. (2003). Generalized autoregressive moving average models. *J. Am. Stat. Assoc.*, 98(461):214–223.
- Bergmeir, C. and Benitez, J. (2012). On the use of cross-validation for time series predictor evaluation. *Inf. Sci.*, 191:192–213.
- Bergmeir, C., Hyndman, R., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Comput. Stat. Data Anal.*, 120:70–83.
- Bourguignon, M. and Vasconcellos, K. (2018). The effects of additive outliers in INAR(1) process and robust estimation. *Stat. Theory Relat. Fields*, pages 206–214.
- Box, G. and Jenkins, G. (1976). *Time series analysis: forecasting and control*. Holden-Day, San Francisco.

- Bustos, O. and Yohai, V. (1986). Robust estimates for ARMA models. *J. Am. Stat. Assoc.*, 81:155–168.
- Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *J. Am. Stat. Assoc.*, 96(455):1022–1030.
- Carroll, R. and Welsh, A. (1986). A note on asymmetry and robustness in linear regression. *Am. Stat.*, 42:285–287.
- Chen, R., C., Tan, J., Cao, J., Song, W., Xu, X., Jiang, C., W., M., Yang, C., Chen, B., Gui, Y., and Kan, H. (2010). Ambient air pollution and hospital admission in Shanghai, China. *J. Hazard. Mater.*, 181(1):234–240.
- Chi, E. (1994). M-estimation in cross-over trials. *Biometrics*, 50(2):486–493.
- Cox, D. (1981). Statistical analysis of time series: some recent developments. *Scand. J. Stat.*, 8:93–115.
- Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65:1254–1261.
- Davis, R. and Dunsmuir, W. (2016). State space models for count time series. In *Handbook of discrete-valued time series*, Chapman & Hall/CRC Handb. Mod. Stat. Methods, pages 121–144. CRC Press, Boca Raton, FL.
- Davis, R., Dunsmuir, W., and Streett, S. (2003). Observation driven models for Poisson counts. *Biometrika*, 90(4):777–790.
- Davis, R., Dunsmuir, W., and Streett, S. (2005). Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodol. Comput. Appl. Probab.*, 7(2):149–159.
- Davis, R., Fokianos, K., Holan, S., and Joe, H. (2021). Count time series: a methodological review. *J. Am. Stat. Assoc.*, 116(535):1533–1547.
- Davis, R., Wang, Y., and Dunsmuir, W. (1999). *Modelling Time Series of Count Data*. In: S. Ghosh (ed) *Asymptotics, Nonparametrics, and Time Series*. CRC Press, New York.
- Dockery, D. and Pope, C. (1996). *Epidemiology of Acute Health Effects: Summary of time series study*. In Richard Wilson and John Spengler, eds., *Particles in our air*. Harvard University Press, Cambridge, MA.
- Dunsmuir, W. and Scott, D. (2015). The glarma package for observation-driven time series regression of counts. *J. Stat. Softw.*, 67(7).
- Elsaied, H. and Fried, R. (2021). On robust estimation of negative binomial INARCH models. *Metron*, 79:137–158.
- Fokianos, K. and Tjostheim, J. (2011). Log-linear Poisson autoregression. *J. Multivar. Anal.*, 102(3):563–578.
- Fox, A. (1972). Outliers in time series. *J. R. Stat. Soc., Ser. B*, 34(3):350–363.
- Froes, C., Camara, V., Landrigan, P., and Claudio, L. (2016). Systematic review of children’s environmental health in Brazil. *Ann. Glob. Health*, 82(1):132–148.

- Hampel, F. (1968). Contributions to the theory of robust estimation. *Ph.D. thesis, University of California, Berkeley.*
- Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions.* New York: John Wiley and Sons, Inc.
- Harvey, A. and Fernandes, C. (1989). Time series models for count or qualitative observations. *J. Bus. Econ. Stat.*, 7(4):407–417.
- Huber, P. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.*, 35(1):73–101.
- Huber, P. (1981). *Robust Statistics.* Wiley, New York.
- Huebnerova, Z. and Michalek, J. (2014). Analysis of daily average PM<sub>10</sub> predictions by generalized linear models in Brno, Czech Republic. *Atmospheric Pollution Research*, 5(3):471–476.
- Junger, W. and De Leon, A. (2015). Imputation of missing data in time series for air pollutants. *Atmos. Environ.*, 102:96–104.
- Karami, S., Karami, M., Roshanaei, G., and Farsan, H. (2017). Association between increased air pollution and mortality from respiratory and cardiac diseases in Tehran: Application of the GLARMA model. *Iran. J. Epidemiol.*, 12(4):36–43.
- Kim, B. and Lee, S. (2017). Robust estimation for zero-inflated Poisson autoregressive models based on density power divergence. *J. Stat. Comput. Simul.*, 87(15):2981–2996.
- Kim, J.-Y., Kim, H.-Y., Park, D., and Chung, Y. (2018). Modelling of fault in RPM using the GLARMA and INGARCH model. *Electron. Lett.*, 54(5):297–299.
- Kitromilidou, S. and Fokianos, K. (2016). Robust estimation methods for a class of log-linear count time series models. *J. Stat. Comput. Simul.*, 86(4):740–755.
- Künsch, H., Stefanski, L., and Carroll, R. (1989). Conditionally unbiased bounded-influence estimation in general regression models with applications to generalized linear models. *J. Am. Stat. Assoc.*, 84:460–466.
- Li, Q., Lian, H., and Zhu, F. (2016). Robust closed-form estimators for the integer-valued GARCH(1,1) model. *Comput. Stat. Data Anal.*, 101:209–225.
- Li, T. (2008). Laplace periodogram for time series analysis. *J. Am. Stat. Assoc.*, 103:757–768.
- Lippmann, M. (2014). Toxicological and epidemiological studies of cardiovascular effects of ambient air fine particulate matter (PM<sub>2.5</sub>) and its chemical components: Coherence and public health implications. *Critical Reviews in Toxicology*, 44(4):299–347.
- Lo, S. and Ronchetti, E. (2009). Robust and accurate inference for generalized linear models. *J. Multivar. Anal.*, 100:2126–2136.
- Lévy-Leduc, C., Bondon, P., and Reisen, V. A. (2022). Robust autocovariance estimation from the frequency domain for univariate stationary time series. *J. Stat. Plan. Inference*, 221:281–298.
- Mallows, C. (1975). On some topics in robustness. *Technical Memorandum, Murray Hill, N.J., Bell Telephone Laboratories.*
- Maronna, R., Martin, R., and Yohai, V. (2006). *Robust Statistics: Theory and Methods.* John Wiley & Sons, Chichester.

- Martin, R. and Yohai, V. (1985). *Robust identification of autoregressive moving average models*. In *Handbook of Statistics 5* (eds E. J. Hannan, P. R. Krishnaiah and M. M. Rao). Elsevier Science Publishers, Amsterdam.
- Muler, N. and Yohai, V. (2008). Robust estimates for GARCH models. *J. Stat. Plan. Inference*, (138):2918–2940.
- Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *J. R. Stat. Soc., Ser. A*, 135(3):370–384.
- Ostro, B., Eskeland, G., Sánchez, J., and Feyzioglu, T. (1999). Air pollution and health effects: a study of medical visits among children in Santiago, Chile. *Environ. Health Perspect.*, 107(1):69–73.
- Ostro, B., Roth, L., Malig, B., and Marty, M. (2009). The effects of fine particle components on respiratory hospital admissions in children. *Environ. Health Perspect.*, 117(3):475–480.
- Peitzsch, E., Pederson, G., Birkeland, K., Hendriks, J., and Fagre, D. (2021). Climate drivers of large magnitude snow avalanche years in the U.S. northern rocky mountains. *Sci. Rep.*, 11(10032).
- Pope, C. and Dockery, D. (2018). Health effects of fine particulate air pollution: lines that connect. *J. Air Waste Manage. Assoc.*, 23(2):307–318.
- Pope, C., Thun, M., Namboodiri, M., Dockery, D., Evans, J., Speizer, F., and Heath, C. (1995). Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am. J. Respir. Crit. Care Med.*, 151:669–674.
- Reisen, V., Levy-Leduc, C., and Taqqu, M. (2017). An M-estimator for the long-memory parameter. *J. Stat. Plan. Inference*, 187:44–55.
- Reisen, V. A., Lévy-Leduc, C., Bondon, P., Cotta, H., Ispány, M., and Filho, P. (2020). An overview of robust spectral estimators. In *Time Series and Cyclostationary process*. Springer series.
- Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. John Wiley and Sons, Inc, New York, USA.
- Ruckstuhl, A. and Welsh, A. (1999). Robust fitting of the binomial model. *CMA, Australian National University, manuscript*.
- Rydberg, T. and Shephard, N. (2003). Dynamics of trade-by-trade price movements:decomposition and models. *J. Financ. Econom.*, 1(1):2–25.
- Schwartz, J. (2000). Harvesting and long-term exposure effects in the relationship between air pollution and mortality. *Am. J. Epidemiol.*, 151(5):440–448.
- Simpson, D., Ruppert, D., and Carroll, R. (1992). On one-step GM estimates and stability of inferences in linear regression. *J. Am. Stat. Assoc.*, 87(418):439–450.
- Solci, C., Reisen, V. A., and Canas, P. (2023). Robust local bootstrap for weakly stationary time series in the presence of additive outliers,. *Stoch. Environ. Res. Risk Assess.*, 37(8):2977–2992.
- Souza, J., Reisen, V., Franco, G., Ispány, M., Bondon, P., and Santos, J. (2018). Generalized additive models with principal component analysis: an application to time series of respiratory disease and air pollution data. *J. R. Stat. Soc., Ser. C*, 67(2):453–480.

- Streett, J., Carroll, R., and Ruppert, D. (1988). A note on computing robust regression estimates via iterative reweighted least squares. *J. Am. Stat. Assoc.*, 42(2):152–154.
- Tukey, J. (1960). *A survey of sampling from contaminated distributions. In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (I. Olkin et al., eds.)*. Stanford Univ. Press.
- Valdora, M. and Yohai, V. (2014). Robust estimators for generalized linear models. *J. Stat. Plan. Inference*, 146:31–48.
- van Buuren, S. and Oudshoorn, K. (2000). Multivariate imputation by chained equations: Mice v1.0 user’s manual. *TNO Prevention and Health*.
- van Buuren, S. and Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.*, 45(3):1–67.
- Weiss, C. H. (2018). *An introduction to discrete-valued time series*. John Wiley & Sons, New Jersey, USA.
- Wu, W. (2007). M-estimation of linear models with dependent errors. *Ann. Stat.*, 35(2):495–521.
- zan Zoest, V., Stein, A., and Hoek, G. (2018). Outlier detection in urban air quality sensor networks. *Water Air Soil Pollut.*, 229(111).
- Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, 44(4):1019–1031.

## 7 Appendix

### 7.1 Covariate contaminated by additive outliers

#### 7.1.1 GLARMA(0,1) model

The GLARMA(0,1) model is defined by (11), where  $\{Z_t\}$  satisfies

$$Z_t = \theta e_{t-1} = \theta(Y_{t-1} - \exp(W_{t-1}) \exp(-W_{t-1}/2)), \quad (16)$$

where we set  $\theta = 0.2$ .

Table 15 presents the parameter estimation results when  $\xi = 0$ . For  $n = 100$  and no outlier, the CMLE of  $\beta_0$  and  $\beta_1$  are close to the true values, while the estimator of  $\theta$  underestimates  $\theta$ . For  $n = 100$  and additive outliers, the CMLE seriously underestimate  $\beta_1$  and  $\theta$ . The mean squared error (MSE) increases in the presence of outliers for all parameters. For  $n = 100$  and no outlier, the MQLE of  $\beta_0$  and  $\beta_1$  are close to the CMLE, and the estimator of  $\theta$  is closer to the true value of  $\theta$ . For  $n = 100$  and additive outliers, contrarily to the CMLE, the MQLE are close to the true values. Similar conclusions are observed for  $n = 1000$ , but the values of the MSE are smaller.

Table 16 displays the parameter estimation results when  $\xi = 0.4$ . Similarly to Table 15, the CMLE in the absence of contamination are close to the true values, except for  $\theta$ , which is underestimated. The CMLE underestimate  $\beta_1$  and  $\theta$  when additive outliers exist. Note that the impact of outliers on the estimation of  $\beta_1$  is more prominent than in Table 15, which suggests that the presence of perturbations in a time-correlated covariate must be carefully treated. Without outliers, the MQLE are close to the true values with MSE slightly larger than the CMLE, except for the estimate of  $\theta$ . In the presence of outliers, the MQLE are close to the true values and have similar MSE as the CMLE in the absence of outliers. The same observations hold for  $n = 100$  and  $n = 1000$ .

Table 15: GLARMA(0,1) model with  $\xi = 0$ 

		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.986	0.0057	1.055	0.0084	0.997	0.0006	1.029	0.0014
	$\hat{\beta}_1$	0.509	0.0036	0.403	0.0139	0.503	0.0003	0.403	0.0095
	$\hat{\theta}$	0.121	0.0096	0.079	0.0171	0.126	0.0057	0.094	0.0114
MQLE	$\hat{\beta}_0$	0.968	0.0069	0.979	0.0066	0.984	0.0009	0.988	0.0008
	$\hat{\beta}_1$	0.539	0.0054	0.514	0.0049	0.533	0.0014	0.514	0.0005
	$\hat{\theta}$	0.168	0.0063	0.158	0.0069	0.175	0.0011	0.171	0.0013

Table 16: GLARMA(0,1) model with  $\xi = 0.4$ 

		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.980	0.0064	1.000	0.0055	0.997	0.0005	1.019	0.0009
	$\hat{\beta}_1$	0.493	0.0044	0.338	0.0281	0.501	0.0003	0.424	0.0060
	$\hat{\theta}$	0.111	0.0117	0.112	0.0124	0.120	0.0065	0.108	0.0087
MQLE	$\hat{\beta}_0$	0.964	0.0078	0.977	0.0066	0.982	0.0009	0.987	0.0007
	$\hat{\beta}_1$	0.532	0.0063	0.479	0.0047	0.532	0.0014	0.517	0.0006
	$\hat{\theta}$	0.161	0.0072	0.160	0.0077	0.168	0.0015	0.166	0.0016

## 7.2 Response variable contaminated by additive outliers

### 7.2.1 GLARMA(0,1) model

We consider model (11) where  $\{Z_t\}$  is defined by (16) with  $\theta = 0.2$ .

Table 17 presents the parameter estimation results when  $\xi = 0$ . For  $n = 100$  and no outlier, the CMLE of  $\beta_0$  and  $\beta_1$  close to the true values, and  $\theta$  is underestimated. With additive outliers on  $(Y_t)$ , all CMLE are impacted:  $\beta_0$  is overestimated,  $\beta_1$  and  $\theta$  are underestimated, and all MSE increase. Without additive outliers on  $\{Y_t\}$ , the MQLE of  $\beta_0$  and  $\beta_1$  are close to the CMLE, and the estimate of  $\theta$  is closer to the true value with a smaller MSE. For  $n = 100$  and additive outliers, contrarily to the CMLE, the MQLE are not affected and provide estimates close to the true parameters. Similar conclusions are observed for  $n = 1000$ .

Table 17: GLARMA(0,1) model with  $\xi = 0$ 

		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.977	0.0071	1.121	0.0194	0.996	0.0006	1.137	0.0193
	$\hat{\beta}_1$	0.526	0.0047	0.388	0.0153	0.504	0.0003	0.468	0.0012
	$\hat{\theta}$	0.120	0.0099	0.009	0.0371	0.125	0.0058	0.031	0.028
MQLE	$\hat{\beta}_0$	0.968	0.0077	0.977	0.0071	0.982	0.0009	0.997	0.0006
	$\hat{\beta}_1$	0.539	0.0053	0.526	0.0047	0.533	0.0014	0.531	0.0013
	$\hat{\theta}$	0.171	0.0073	0.168	0.0060	0.172	0.0012	0.165	0.0017

Table 18 presents the parameter estimation results when  $\xi = 0.4$ . For  $n = 100$ , in the absence of

contamination on  $\{Y_t\}$ , the CMLE of  $\beta_0$  and  $\beta_1$  are close to the true values, and  $\theta$  is underestimated. For  $n = 1000$ , the CMLE of  $\beta_0$  and  $\beta_1$  are very close to the true values and  $\theta$  is still underestimated. In the presence of outliers, all CMLE are affected for  $n = 100$  and  $n = 1000$ . For both sample sizes, without outliers, the MQLE gives good results. In the presence of outliers, the MQLE give almost the same results as the CMLE in the absence of outliers, with similar MSE.

Table 18: GLARMA(0,1) model with  $\xi = 0.4$

		$n = 100$				$n = 1000$			
		no outlier		with outliers		no outlier		with outliers	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
CMLE	$\hat{\beta}_0$	0.993	0.0056	1.117	0.0184	0.998	0.0005	1.303	0.0921
	$\hat{\beta}_1$	0.447	0.0059	0.396	0.0134	0.502	0.0003	0.318	0.0331
	$\hat{\theta}$	0.116	0.0104	0.013	0.0351	0.125	0.0058	-0.005	0.0425
MQLE	$\hat{\beta}_0$	0.982	0.0064	0.992	0.0066	0.983	0.0008	1.043	0.0024
	$\hat{\beta}_1$	0.505	0.0040	0.492	0.0037	0.536	0.0017	0.507	0.0004
	$\hat{\theta}$	0.167	0.0061	0.154	0.0067	0.172	0.0012	0.141	0.0039