Robust estimate for counting time series using GLARMA models

Ana Júlia A. Câmara[1*], Valdério A. Reisen[2] and Pascal Bondon[3]

*Corresponding author
[1]Federal University of Minas Gerais, Brazil
[2]Federal University of Espírito Santo, Brazil
[3]Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, France
E-mails: anajulia.camara@gmail.com / valderio.reisen@ufes.br / pascal.bondon@l2s.centralesupelec.fr

Abstract. The generalized linear autoregressive moving average (GLARMA) model has been used in epidemiological studies to evaluate the impact of air pollutants on human health, as frequently, the response variable is a nonnegative integer-valued time series. The relative risk (RR) measure commonly quantifies these health effects. Due to the nature of the data, a robust approach for the GLARMA model is proposed based on the robustification of the quasi-likelihood function. In this method, outlying observations are bounded separately by weight functions on covariates and by the Huber loss function on the response variable. A numerical study was realized to evaluate the performance of the proposed methodology for distinct sample sizes. In real data analysis, the impact of the particulate pollutant PM_{10} in the monthly number of deaths in Vitoria, Brazil, was investigated, showing that the parameter estimates involving the robust method are more reliable than the classic.

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

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 populations' quality of life. Ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), 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) offer 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), Froes et al. (2016) among others).

Epidemiological data are frequently treated as counting 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 the economy, medicine, agriculture, sports, among others. Examples are the monthly number of hospital admissions caused by a disease, the number of car accidents in a city, and the number of 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)), a procedure that expands the possibilities for the distribution of the response variable, which can assume distributions belonging to the exponential family, e.g., Normal, Poisson, Gamma, Negative Binomial, etc. In addition, the relation between the mean of the dependent variable (μ) and the linear predictor (η) 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 generalized linear autoregressive moving average models (GLARMA). Fokianos and Tjosthein (2011) proposed log-linear models for time series. Other procedures can be found in Davis et al. (2021), which realized an overview of methodologies for count time series. Although many methods have been developed in the field, they all present limitations that contribute to the non-development of a unified theory. Despite this, Davis et al. (2021) addresses that the GLARMA family is "one of the most flexible and easily fit count models that balance parameter and observation-driven models". In this methodology, an ARMA structure (Box and Jenkins (1976)) is added to the GLM, allowing the modeling of correlated observations from the exponential family. Even though GLARMA 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.

Studying the statistical association between air pollutants and health effects is complex and must be computed with caution independently of the statistical regression and time series models used. In the epidemiological context, the response variable is usually time correlated, and this should be taken into account. In addition, the dynamic of the response variable and, therefore, the statistical functions that measure the impact of the pollutants on health can not be fully explained by the response variable itself or by only one contaminant since the population under the study is exposed to a complex mixture of pollutants and chemical compounds. Many authors have been ignoring 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 profoundly impacted, leading to false conclusions regarding the population's health risk in generalized additive models. Finally, 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 characteristics of the data, like mean, variance, and correlation. In addition, many authors have been verifying that the presence of 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 is composed of 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. Ledolter (1989) showed that the ARMA models could be substantially affected by additive outliers. Chang et al. (1988) and Chen and Liu (1993) verified that the presence of additive outliers could bias the parameter estimates of the ARMA model. A similar conclusion was obtained by Reisen et al. (2017) and Sarnaglia et al. (2021) for fractionally integrated and periodic ARMA processes. Fokianos and Tjosthein (2011) verified that the maximum likelihood estimator in log-linear Poisson models is highly affected by additive outliers.

The nonrobustness 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), Ruckstuhl and Welsh (1999), and others). Due to this, 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 the most relevant 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' (Mallows (1975)). In this method, outlying observations are bounded separately by weight functions on covariates and by a loss function on the response variable. 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. At the same time, in the presence of additive outliers, the MQLE provided more reliable results. Actually, 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 the best of our knowledge, robustified proposals for the GLARMA using M-estimators are still not explored in the literature. This paper aims to fill this gap. Due to the limitations regarding the asymptotic properties of the GLARMA model, we considered the development of asymptotic theory for the proposed robust approach beyond the scope of this work. In fact, Davis et al. (2021) claim that after all these years theoretical properties for the classic GLARMA model were only established for very restrictive special cases. However, although a general asymptotic theory has not yet been developed, the simulation study showed that asymptotic results corroborate that the estimators are consistent.

A Monte Carlo study was realized to evaluate the impact of additive outliers in the response variable and covariates, considering the classic GLARMA (proposed by Davis et al. (2003)) and the robust proposal under distinct scenarios and sample sizes. Additionally, real data analysis was realized to study the effect of Particulate Material (PM_{10}) on the deaths caused by respiratory diseases in Vitoria, Brazil.

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

2 The generalized linear autoregressive moving average model

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

Let $\{Y_t\} := \{Y_t\}_{t\in\mathbb{Z}}$ be the observations on the response series, $\boldsymbol{X}_t = (X_{1,t}, X_{2,t}, ..., X_{k,t})^T$ the vector of k covariates observed for $t = 1, ..., n$, and $\mathcal{F}_{t-1} = \sigma\{Y_s, s \leq t-1\}$ the process history. The observation process Y_t conditioned on $\mathcal{F}_{t-1}$ is assumed exponentially distributed with density

$$f(Y_t|W_t) = \exp\{Y_t W_t - a_t b(W_t) + c_t\}, \tag{1}$$

where $\{W_t\} := \{W_t\}_{t\in\mathbb{Z}}$ is the canonical parameter that summarizes the information in $\mathcal{F}_{t-1}$, and a_t and c_t are sequences of constants (for more, see Dunsmuir (2015) and Davis et al. (2021)). The conditional mean and variance of Y_t are $\mu_t = \mathrm{E}(Y_t|\mathcal{F}_{t-1})$ and $\sigma_t^2 = \mathrm{Var}(Y_t|\mathcal{F}_{t-1})$, respectively.

The specification of $W_t = \ln(\mu_t)$ is given by

$$W_t = \boldsymbol{X}_t^T \boldsymbol{\beta} + Z_t, \tag{2}$$

where $\boldsymbol{\beta}$ is a $(k+1)$ x 1 vector of unknown coefficients, and the noise process $\{Z_t\}_{t\in\mathbb{Z}}$, which induces a serial dependence on the observation, is given by

$$Z_t = \sum_{i=1}^{\infty} \gamma_i e_{t-i}. \tag{3}$$

The parameters γ_i's are the coefficients in the power series expansion

$$\sum_{i=1}^{\infty} \gamma_i B^i = \frac{\theta(B)}{\phi(B)} - 1, \tag{4}$$

where the autoregressive and moving average components $\phi(B) = (1 - \phi_1 B - ... - \phi_p B^p)$ and $\theta(B) = (1 + \theta_1 B + ... + \theta_q B^q)$ are polynomials with no common zeroes and have all their zeros outside the unit circle. The parameter vector γ is formed by ϕ's and θ's, and B is the backshift operator of the form $B^k(Z_t) = Z_{t-k}$. From (3) and (4) Z_t can be calculated recursively with the difference equation

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

The predictive residuals $\{e_t\}_{t\in\mathbb{Z}}$, in (3) are given by

$$e_t = \frac{Y_t - \mu_t}{\nu_t}, \tag{6}$$

where $\nu_t = \sigma_t$ for Pearson residuals. From (6), $\mathrm{E}(e_t|\mathcal{F}_{t-1}) = (\mathrm{E}(Y_t|\mathcal{F}_{t-1}) - \mu_t)/\nu_t = 0$. Under the initial conditions $e_s = 0$ and $Y_s = 0$, for $s \leq 0$, let $\mathcal{F}^e_{t-1} = \sigma(e_s, s \leq t-1)$. Equation (6) implies $\mathcal{F}^e_{t-1} \subset \mathcal{F}_{t-1}$, therefore

$$\mathrm{E}(e_t|\mathcal{F}^e_{t-1}) = \mathrm{E}[\mathrm{E}(e_t|\mathcal{F}_{t-1})|\mathcal{F}^e_{t-1}] = 0,$$

which means that $\{e_t\}$ are martingale differences, with Cov(e_t, e_s)=0, for $t \neq s$. For Pearson residuals,

$$\mathrm{Var}(e_t) = \mathrm{E}(e_t^2) = \mathrm{E}[\mathrm{E}(e_t^2|\mathcal{F}_{t-1})] = \mathrm{E}\left[\mathrm{E}\left(\tfrac{Y_t-\mu_t}{\sigma}\right)^2 \,\middle|\, \mathcal{F}_{t-1}\right] = \mathrm{E}\left[\tfrac{\mathrm{E}(Y_t-\mu_t|\mathcal{F}_{t-1})^2}{\sigma^2}\right] = 1,$$

i.e. $\{e_t\}$ are weakly stationary white noise.

Considering n successive observations $y_1, y_2, ..., y_n$, the likelihood is constructed as the product of conditional densities of $\{Y_t\}$ given $\mathcal{F}_{t-1}$, corresponding to the following log-likelihood

$$L(\boldsymbol{\delta}) = \textstyle\sum_{t=1}^n \left\{Y_t W_t(\boldsymbol{\delta}) - a_t b(W_t(\boldsymbol{\delta})) + c_t\right\},$$

where $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T, \boldsymbol{\theta}^T)^T$ is the parameter vector.

For the particular case of Poisson distribution, where $Y_t|\mathcal{F}_{t-1} \sim$ Poisson(μ_t) with $\mu_t = e^{W_t}$, the log-likelihood is given by

$$L(\boldsymbol{\delta}) = \sum_{t=1}^{n} \left\{y_t W_t(\boldsymbol{\delta}) - e^{W_t(\boldsymbol{\delta})} - \log(y_t!)\right\}. \tag{7}$$

The log-likelihood can be maximized using Newton-Raphson iterations or Fisher scoring procedure from suitable initial values by computing the first and second derivatives of the likelihood. According to Davis et al. (2005) the first derivative for (7) is given by

$$\frac{\partial L(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = \sum_{t=1}^{n} (y_t - \mu_t) \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}, \tag{8}$$

and the second derivative is

$$\frac{\partial L^2(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} = \sum_{t=1}^{n} (y_t - \mu_t) \frac{\partial^2 W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} - \sum_{t=1}^{n} \mu_t \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^T}. \tag{9}$$

$E(y_t - \mu_t | \mathcal{F}_{t-1}) = 0$ at the true value of $\boldsymbol{\delta}$, which implies that the first summation in (9) is zero. This motivates the Fisher-scoring approximation based only on the first derivatives

$$\frac{\partial L^2(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} = - \sum_{t=1}^{n} \mu_t \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^T}. \tag{10}$$

Note that although $E(D_{NR}(\boldsymbol{\delta})) = E(D_{FS}(\boldsymbol{\delta}))$, these expectations can not be calculated in closed form. Thus, the maximum likelihood estimation $\hat{\boldsymbol{\delta}}$ can be computed using the Newton-Raphson iterations (based on equations (8) and (9)) or the Fisher scoring approximations (equations (8) and (10)).

3 Robust estimation

Given a parametric model F_δ, a general M-estimate (Huber (1981)) of $\delta \in \Delta$, say $\hat{\delta}$, is defined as a solution of

$$\arg \min_{\delta} \sum_{i=1}^{n} \rho(\xi_i, \delta), \tag{11}$$

where $\rho(\cdot)$ is the loss function. Supose the solution of (11) is an interior point, if $\rho(\cdot)$ is differentiable with respect to δ, where $\psi(\cdot) = \rho'(\cdot)$, then $\hat{\delta}$ is the solution of the estimating equations

$$\sum_{i=1}^{n} \psi(\xi_i, \delta) = 0. \tag{12}$$

There are several candidates for ρ. However, the Huber loss function, proposed by Huber (1964), and Tukey's biweight, given by (Beaton and Tukey (1974)) are the most used. Another loss functions can be found in Hampel (1974), Andrews (1974), Dennis and Welsch (1978), and Maronna et al. (2006). According to Huber (1981), $\rho(\cdot)$ must to satisfy the following assumptions:

A1) $\rho(0) = 0$;

A2) $\rho(\xi) = \rho(-\xi) \forall \xi \in \mathbb{R}$, i.e. $\rho(\xi)$ is a symmetric function;

A3) $0 \leq \xi \leq \xi^* \Rightarrow \rho(\xi) \leq \rho(\xi^*), \forall (\xi, \xi^*) \in \mathbb{R}^2$;

A4) $\psi(\cdot)$ is bounded;

A5) $\rho(\cdot)$ has a second derivative almost everywhere.

Here we will focus on the Huber loss function

$$\rho_H(\xi) = \begin{cases} \frac{1}{2}\xi^2, & |\xi| \leq c \\ c|\xi| - \frac{1}{2}c^2, & |\xi| > c. \end{cases} \tag{13}$$

Its derivative, ψ-function, is given by

$$\psi_H(\xi) = \begin{cases} \xi, & |\xi| \leq c \\ c\text{sign}\,(\xi)\,, & |\xi| > c, \end{cases} \tag{14}$$

where the constant c must be prespecified and regulates the amount of robustness. This parameter regulates the trade-off between the efficiency and robustness of the estimators. Good choices for the constant value are in the range between 1 and 2. According to (Huber (1964)) $c = 1.345$ provides 90% efficiency when the data is normally distributed. 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) and Chi (1994)). The choice of c should reflect the proportion of outliers in the data. Moreover, this value must be adjusted according to the data distribution.

3.1 Robust estimation for GLARMA models

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

The MQLE for GLARMA family, denoted by $\hat{\delta}_{MQLE}$, is the solution of the estimating equation

$$S_n(\delta) = \sum_{t=1}^{n} \left[\nu(Y_t, \mu_t) w(\mathbf{X}_t) \mu_t' - \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}\left(\nu(Y_t, \mu_t) | \mathcal{F}_{t-1}\right) w(\mathbf{X}_t) \mu_t' \right] = 0. \tag{15}$$

The $\hat{\delta}_{MQLE}$ is an M-estimator (Huber (1981); Hampel et al. (1986)) characterized by the score function $S_n(\delta) = \sum_{t=1}^{n} [\nu(Y_t, \mu_t) w(\mathbf{X}_t) \mu_t' - a(\boldsymbol{\delta})]$, where $a(\boldsymbol{\delta}) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}(\nu(Y_t, \mu_t) w(\mathbf{X}_t) \mu_t' | \mathcal{F}_{t-1})$ is a bias correction used to ensure Fisher's consistency. Additionally, function $\nu(\cdot, \cdot)$ is chosen to control deviations on Y-space and leverage points on $\mathbf{X}$-space are down-weighted by $w(\cdot)$.

Künsch (1984) extended the definition of Influence Function (IF) of Hampel (1974) to time series for stationary process. Therefore, considering a cumulative distribution function F, the $\mathrm{IF}(Y_t; \boldsymbol{\psi}, F) = M(\boldsymbol{\psi}, F)^{-1} S_n(\delta)$, where $M(\boldsymbol{\psi}, F)^{-1} = -\mathrm{E}\left[\frac{\partial}{\partial \beta} S_n(\delta)\right]$, for more, see Maronna et al. (2006). Choosing a bounded function S_n leads to limits on the influence function, which ensures the robustness of the estimator. Thus, bounded functions $\nu(\cdot, \cdot)$ and $w(\cdot)$ must be chosen to restrict outlying values on the response variable and covariates, respectively.

Let $\nu(Y_t, \mu_t) = \psi_H(r_t)\frac{1}{Var(Y_t)^{1/2}}$, where, ψ_H is the Huber loss function defined in (14), $r_t = \frac{Y_t - \mu_t}{Var(Y_t)^{1/2}}$, are the Pearson residuals and $\mu_t' = \mu_t \frac{\partial W_t(\boldsymbol{\delta})}{\partial \delta}, t = 1, ..., n$. Replace it on (15), then $\hat{\delta}_{MQLE}$ of the GLARMA Poisson model is the solution of the following equation

$$S_n(\delta) = \sum_{t=1}^{n} \left[\frac{\psi_H(r_t)}{Var(Y_t)^{1/2}} w(\mathbf{X}_t)\mu_t \frac{\partial W_t(\boldsymbol{\delta})}{\partial \delta} - \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left(\frac{\psi_H(r_t)}{Var(Y_t)^{1/2}} \,\middle|\, \mathcal{F}_{t-1} \right) w(\mathbf{X}_t)\mu_t \frac{\partial W_t(\boldsymbol{\delta})}{\partial \delta} \right] = 0, \tag{16}$$

where $Var(Y_t) = \mu_t$ and

$$\begin{aligned}\mathrm{E}\left(\psi_H(r_t) \,\middle|\, \mathcal{F}_{t-1} \right) &= c\left\{ P\left(Y_t \geq j_2 + 1 | \mathcal{F}_{t-1}\right) - P\left(Y_t \leq j_1 | \mathcal{F}_{t-1}\right) \right\} \\ &\quad + \mu_t^{1/2}\left\{ P(Y_t = j_1 | \mathcal{F}_{t-1}) - P(Y_t = j_2 | \mathcal{F}_{t-1}) \right\},\end{aligned}$$

with $j_1 = \left\lfloor \mu_t - c\mu_t^{1/2} \right\rfloor$ and $j_2 = \left\lfloor \mu_t + c\mu_t^{1/2} \right\rfloor$.

A common choice for the sequence of weights $w(\mathbf{X}_t), t = 1, ..., n$, in (16) is $w(\mathbf{X}_t) = \sqrt{1 - h_{tt}}$, where h_{tt} is the tth diagonal element of the hat matrix $H = \mathbf{X}_t(\mathbf{X}_t^T\mathbf{X}_t)^{-1}\mathbf{X}_t^T$ (see Cantoni and Ronchetti (2001)). However, the hat matrix does not have breakdown points, i.e., the estimates are not reasonable if large atypical values contaminate the data. More sophisticated methods can be found in the literature based on the inverse of robust Mahalanobis distance.

Let μ and $\sum$ be the location parameter and the covariance matrix of $\mathbf{X}_t$, respectively. The squared Mahalanobis distance of each observation along a row in $\mathbf{X}_t$ from μ with respect to $\sum$ is

$$d_{\mu,\sum}(\mathbf{X}_t)^2 = (\mathbf{X}_t - \hat{\mu})^T \hat{\sum}^{-1} (\mathbf{X}_t - \hat{\mu}).$$

To robustify the Mahalanobis distance, the location parameters and the covariance matrix can be estimated using the minimum covariance determinant algorithm, the fast MCD (see Rousseeuw (1984), page 877 and Rousseeuw (1985) for more details). In this procedure, h observations (out n) are chosen whose classical covariance matrix presents the lowest determinant. Then the MCD estimate of location ($\hat{\mu}_{(\mathrm{MCD})}$) is the average of the h points, and their covariance matrix is the MCD estimate scatter ($\hat{\sum}_{(\mathrm{MCD})}$). In this paper, we use the weight function $w(\cdot)$ based on the MCD estimates. It is given by

$$w(\mathbf{X}_t) = \min\left[1, \left\{ \frac{b}{(Y_t - \hat{\mu}_{(\mathrm{MCD})})^T \hat{\sum}_{(\mathrm{MCD})}^{-1} (Y_t - \hat{\mu}_{(\mathrm{MCD})})} \right\}^{\alpha/2} \right], \tag{17}$$

where α and b are tuning constants. Simpson et al. (1992) evaluate some values for the constant α 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 $\gamma = 0.1$ and 0.05.

Solve equation (15) corresponds to minimize the following equation

$$Q(\boldsymbol{\delta}) = \sum_{t=1}^{n} Q_M(\delta), \tag{18}$$

where $Q_M(\delta)$ is given by

$$Q_M(\boldsymbol{\delta}) = \int_{\tilde{s}}^{\mu_t} \nu(Y_t,u)w(\mathbf{X}_t)du - \tfrac{1}{n}\sum_{j=1}^{n} \int_{\tilde{u}}^{\mu_j} \mathrm{E}[\nu(Y_j,u)w(\mathbf{X}_j)|\mathcal{F}_{t-1}]du,$$

with $\tilde{s}$ and $\tilde{u}$ defined such as $\nu(Y_t,\tilde{s}) = 0$ and $\mathrm{E}[\nu(Y_t,\tilde{u})] = 0$ (see Cantoni and Ronchetti (2001) for more details).

Parameter estimates can be obtained using Newton-Raphson or Fisher-scoring approximations. The first derivative of $Q(\boldsymbol{\delta})$ is

$$\frac{\partial Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = S_n(\boldsymbol{\delta}). \tag{19}$$

For $r_t \leq c$ the second derivative is

$$\frac{\partial^2 Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} = \sum_{t=1}^{n} \left[(Y_t - \mu_t) \frac{\partial^2 W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} w(\mathbf{X}_t) - \mu_t \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^T} w(\mathbf{X}_t) - a'(\boldsymbol{\delta}) \right]. \tag{20}$$

where

$$a(\boldsymbol{\delta}) = \tfrac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left[\psi_H \left(\tfrac{Y_t - \mu_t}{\mu_t^{1/2}} \right) w(\mathbf{X}_t) \mu_t^{1/2} \tfrac{\partial W_t(\boldsymbol{\delta})}{\partial \delta} \;\middle|\; \mathcal{F}_{t-1} \right].$$

For $r_t > c$ the second derivative of $Q(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ is

$$\frac{\partial^2 Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} = \sum_{t=1}^{n} \left\{ c\mathrm{sign}(r_t) w(\mathbf{X}_t) \left[\frac{1}{2} \mu_t^{1/2} \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^T} + \mu_t^{1/2} \frac{\partial^2 W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \right] - a'(\boldsymbol{\delta}) \right\}. \tag{21}$$

At the true parameter of $\boldsymbol{\delta}$, $E[(y_t - \mu_t)|\mathcal{F}_{t-1}] = 0$, the expected value of the first summation in (20) is zero, which motivates the Fisher Scoring approximation:

$$\frac{\partial^2 Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} = \sum_{t=1}^{n} \left[-\mu_t \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} \frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^T} w(\mathbf{X}_t) - a'(\boldsymbol{\delta}) \right]. \tag{22}$$

Details about the computation of the derivative $a'(\boldsymbol{\delta}) = \frac{\partial a(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}$ and the second derivative of $Q(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ can be found in Appendix 1.

4 Monte Carlo study

A simulation study was conducted to evaluate the performance of the robust estimation for the GLARMA Poisson model proposed in 3.1. The model is given by

$$Y_t|\mathcal{F}_{t-1} \sim \mathrm{Poisson}(\mu_t) \tag{23}$$

$$\ln(\mu_t) = \beta_0 + \beta_1 X_{1,t} + Z_t. \tag{24}$$

Two scenarios were considered to the regressor variable; $(X_{1,t})$ is an independent $N(0,1)$ random variable, and $(X_{1,t})$ is an autoregressive process of order 1. We set $\beta_0 = 1$ and $\beta_1 = 0.5$. The Monte Carlo simulations were repeated 1000 times with sample sizes equal to $n = 100$ and $n = 1000$. The choice of the *tuning parameter* for the Huber function was $c = 1.345$. However, the cross-validation procedure applied to time series, using blocks, is also an option to choose the value of this constant (see Bergmeir and Benitez (2012) and Bergmeir et al. (2018)). Both options were considered in the numerical simulations and provided similar results.

4.1 Covariate contaminated by additive outliers

Additive outliers perturbed the covariate $(X_{1,t})$. The contaminated version of $X_{1,t}$ is defined by $X^*_{1,t} = X_{1,t} + \omega\varphi_t$, where $\omega = 5$ is the magnitude of the outlier which impacts $X_{1,t}$ and φ_t indicates the presence or not of this outlier and its sign at time t, i.e., $\varphi_t = 0$ with probability $1 - \varphi$, $\varphi_t = 1$ with probability $\varphi/2$, and $\varphi_t = -1$ with probability $\varphi/2$, where $\varphi = 0.01$

Once Davis et al. (2003) only presented formal properties for the simplest case, where the time correlation structure is moving average, we will first show the scenario considering the GLARMA(0,1) model and then extend the simulations for the GLARMA(1,0) model.

4.1.1 Scenario 1: Moving average process - GLARMA(0,1)

The GLARMA(0,1) model is defined as equations (23) and (24), where $\{Z_t\}$ is a moving average process of order 1, defined as $Z_t = \theta(Y_{t-1} - e^{\eta_{t-1}})e^{-\lambda\eta_{t-1}}$ with $\theta = 0.2$ and $\lambda = 0.5$, which corresponds to Pearson residuals.

Table 1 presents the parameter estimation considering $X_{1,t}$ as an independent random vector in time, following a distribution Normal(0,1). For $n = 100$, in the classic procedure, without outliers, the mean of $\hat{\beta}_0$ and $\hat{\beta}_1$ was close to the real values, while $\hat{\theta}$ was underestimated. The classic approach in the presence of additive outliers was impacted in the mean of $\hat{\beta}_1$ and $\hat{\theta}$, with both parameters being underestimated. The mean squared error (MSE) increased in the presence of outliers for all parameters in the study. The robust approach without any outlier presented parameter estimations similar to the classic in the same conditions. However, the mean of $\hat{\theta}$ was closer to the real value of θ, and consequently, the MSE was smaller than the error observed in the classic method. Finally, the proposed robust methodology was applied to contaminated data. The results showed that, differently from the classic GLARMA, the mean of the estimates was not affected, with values close to the real ones. The MSE was not impacted as well. Similar conclusions were observed for $n = 1000$, but the values of the MSE were smaller.

In Table 2, $X_{1,t} \sim$ AR(1), with the autoregressive parameter assuming value 0.4. Similarly to Table 1, the classic GLARMA in the absence of contamination presented parameters estimates closer to the real values, except for $\hat{\theta}$, which was underestimated. In the presence of additive outliers, $\hat{\beta}_1$ and $\hat{\theta}$ were again affected. It is important to note

Table 1: Parameter estimation - $X_{1,t} \sim$ N$(0,1)$ - GLARMA(0,1)

		n=100				n=1000			
		no outlier		with outlier		no outlier		with outlier	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
Classic	$\hat{\beta}_0$	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
Robust	$\hat{\beta}_0$	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

that the impact in $\hat{\beta}_1$ was more prominent in this case, which suggests that for covariates with time correlation structure, the presence of perturbation in the data must be carefully treated. For the robust approach, the mean of parameters in the study was close to the actual values, with values of MSE slightly bigger than the classic method, except for $\hat{\theta}$. The robust GLARMA applied to the series contaminated by additive outliers provided parameter estimates close to the real ones and MSE values comparable to the results of the classic method in the absence of outliers. The same conclusions were observed for $n = 100$ and $n = 1000$.

Table 2: Parameter estimation - $X_{1,t} \sim$ AR(1) - GLARMA(0,1)

		n=100				n=1000			
		no outlier		with outlier		no outlier		with outlier	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
Classic	$\hat{\beta}_0$	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
Robust	$\hat{\beta}_0$	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

4.1.2 Scenario 2: Autoregressive process - GLARMA(1,0)

For the GLARMA(1,0) model, defined by equations (23) and (24), $\{Z_t\}$ is an autoregressive process of order 1, where $Z_t = \phi[Z_{t-1} + (Y_{t-1} - e^{\eta_{t-1}})e^{-\lambda\eta_{t-1}}]$, with $\phi = 0.2$, and $\lambda = 0.5$.

Table 3 presents the parameter estimation for $X_{1,t} \sim N(0,1)$. For both sample sizes, in the classic procedure, without perturbations, the mean of $\hat{\beta}_0$ and $\hat{\beta}_1$ was close to the real values, while $\hat{\phi}$ was underestimated. The parameter estimates of the classic approach in the presence of additive outliers were impacted in the mean and MSE of $\hat{\beta}_1$ and $\hat{\phi}$. Both parameters were underestimated. The robust proposal without contamination presented parameter estimations close to the classic in the same conditions, except for $\hat{\phi}$, which was closer to the actual value of ϕ, and displayed an MSE smaller than that observed in the classic method. When the proposed robust GLARMA model was applied to contaminated data, the mean of the estimates was not affected, with values close to the real ones. The MSE has practically not changed compared to the scenario without additive outliers.

Table 3: Parameter estimation - $X_{1,t} \sim \mathrm{N}(1,0)$ - GLARMA(1,0)

		n=100				n=1000			
		no outlier		with outlier		no outlier		with outlier	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
Classic	$\hat{\beta}_0$	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
Robust	$\hat{\beta}_0$	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 for $X_{1,t} \sim \mathrm{AR}(1)$. As observed in Table 3, the classical method in clear data presented parameter estimates closer to the real values, except for $\hat{\phi}$, which was underestimated. In the presence of additive outliers $\hat{\beta}_1$ and $\hat{\phi}$ were affected. Note that $\hat{\beta}_1 = 0.299$, while the real parameter value is 0.50, a significant reduction. For the robust approach, the mean of parameters in the study was close to the actual values, with values of MSE slightly bigger than the classic method, except for $\hat{\phi}$. The robust GLARMA applied to the series contaminated by additive outliers provided parameter estimates close to the real ones and MSE values similar to the classic method in the absence of outliers. Similar results were observed for $n = 100$ and $n = 1000$.

Table 4: Parameter estimation - $X_{1,t} \sim \mathrm{AR}(1)$ - GLARMA(1,0)

		n=100				n=1000			
		no outlier		with outlier		no outlier		with outlier	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
Classic	$\hat{\beta}_0$	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
Robust	$\hat{\beta}_0$	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

4.2 Response variables Y_t contaminated by additive outliers

The effect of additive outliers in count time series was evaluated considering GLARMA(0,1) and GLARMA(1,0) models under scenarios where covariate is an independent random variable and $X_{1,t} \sim AR(1)$. 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 φ_t indicates the presence or not of this outlier at time t, i.e., $\varphi_t = 1$ with probability φ, and $\varphi_t = 0$ with probability $1 - \varphi$, where $\varphi = 0.01$.

4.2.1 Scenario 3: Moving average process - GLARMA(0,1)

GLARMA(0,1) model is defined according equations (23) and (24), where process $\{Z_t\}$ is a moving average of order 1, given by $Z_t = \theta(Y_{t-1} - e^{\eta_{t-1}})e^{-\lambda\eta_{t-1}}$ with $\theta = 0.2$ and $\lambda = 0.5$.

Table 5 presents the parameter estimation for $X_{1,t} \sim N(0,1)$. Under the classical procedure, with clear data, the mean of $\hat{\beta}_0$ and $\hat{\beta}_1$ was close to the real values, and $\hat{\theta}$ was underestimated. However, all parameters were impacted by additive outliers on the count response variable. $\hat{\beta}_0$ was overestimated, while $\hat{\beta}_1$ and $\hat{\theta}$ were underestimated. The MSE increased in the presence of outliers for all parameters in the study. The robust approach without additive outliers on $\{Y_t\}$ presented parameter estimates similar to the classic in the same conditions. But, the mean of $\hat{\theta}$ was closer to the real value of θ, while its MSE was smaller than the in the classic method. The robust methodology applied to contaminated data showed that the mean of the estimates was not affected, with values close to the real ones. The MSE slightly decreased. Similar conclusions were observed for $n = 1000$.

Table 5: Parameter estimation - $X_{1,t} \sim \mathrm{N}(0,1)$ - GLARMA(0,1)

		n=100				n=1000			
		no outlier		with outlier		no outlier		with outlier	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
Classic	$\hat{\beta}_0$	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
Robust	$\hat{\beta}_0$	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 6 presents the parameter estimates for $X_{1,t} \sim \mathrm{AR}(1)$. The classic GLARMA(0,1), for $n = 100$, in the absence of contamination on $\{Y_t\}$ presented estimates closer to the real values, only for $\hat{\beta}_0$. $\hat{\beta}_1$ and $\hat{\theta}$ were underestimated. For $n = 1000$, only $\hat{\theta}$ was underestimated, while $\hat{\beta}_0$ and $\hat{\beta}_1$ were close to actual values. In the presence of outliers, all parameters were affected. Note that the impact for $n = 1000$ was more prominent in this case. For both sample sizes, in the robust approach, without additive outliers, the mean of parameters in the study was close to the real ones. In the presence of perturbations on the response variable, the robust approach provided parameter estimates close to the true values and MSE measures comparable to the results of the classic method in the absence of outliers.

Table 6: Parameter estimation - $X_{1,t} \sim$ AR(1) - GLARMA(0,1)

		n=100				n=1000			
		no outlier		with outlier		no outlier		with outlier	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
Classic	$\hat{\beta}_0$	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
Robust	$\hat{\beta}_0$	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

4.2.2 Scenario 4: Autoregressive process - GLARMA(1,0)

The GLARMA(1,0) model is defined by equations (23) and (24), with the autoregressive process of order 1 $\{Z_t\}$ given by $Z_t = \phi[Z_{t-1} + (Y_{t-1} - e^{\eta_{t-1}})e^{-\lambda\eta_{t-1}}]$, with $\phi = 0.2$, and $\lambda = 0.5$.

Table 7 presents the parameter estimation for $X_{1,t} \sim N(0,1)$. In the classic procedure, without additive outliers, the mean of $\hat{\beta}_0$ and $\hat{\beta}_1$ was close to the real values, while $\hat{\phi}$ was underestimated. All the parameter estimates were impacted in the presence of additive outliers on $\{Y_t\}$. The MSE increased in the presence of outliers for parameters $\hat{\beta}_0$ and $\hat{\beta}_1$, for $n = 100$. For $n = 1000$, the MSE increased for all parameters. The robust approach without outlier presented parameter estimations close to the classic in the same conditions, except for $\hat{\phi}$, which was more relative to the true value of ϕ. As expected, for this parameter, the MSE was smaller than that observed in the classic method. Applying the robust procedure to the contaminated data, the mean of the estimates was not affected, with values close to the real ones.

Table 7: Parameter estimation - $X_{1,t} \sim \mathrm{N}(0,1)$ - GLARMA(1,0)

		n=100				n=1000			
		no outlier		with outlier		no outlier		with outlier	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
Classic	$\hat{\beta}_0$	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
Robust	$\hat{\beta}_0$	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 8 presents the parameter estimation for $X_{1,t} \sim$ AR(1). As observed in Table 7, the classical method, without perturbation, presented parameter estimates closer to the real values, except for $\hat{\phi}$, which was underestimated. In the presence of additive outliers, all parameters were affected. For the robust approach, the mean of parameters in the study was close to the actual values, with values of MSE slightly bigger than the classic method, except for $\hat{\phi}$. The robust GLARMA applied to response variables perturbed by additive outliers provided parameter estimates close to the real ones and MSE values comparable to the classic method in the absence of outliers. Similar results were observed for $n = 100$ and $n = 1000$.

Table 8: Parameter estimation - $X_{1,t} \sim$ AR(1) - GLARMA(1,0)

		n=100				n=1000			
		no outlier		with outlier		no outlier		with outlier	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
Classic	$\hat{\beta}_0$	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
Robust	$\hat{\beta}_0$	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

The empirical study showed that the classic GLARMA is impacted by additive outliers, independently of the perturbation affecting the covariate or the response variable. The proposed robust GLARMA approach provides similar results to the classic in the absence of additive outliers. In contaminated data, the robust approach was superior, providing parameter estimates closer to the true values with small MSE measures. In all scenarios analyzed, the MSEs for $n = 1000$ were smaller than $n = 100$, independently of the methodology applied.

5 Real data analysis

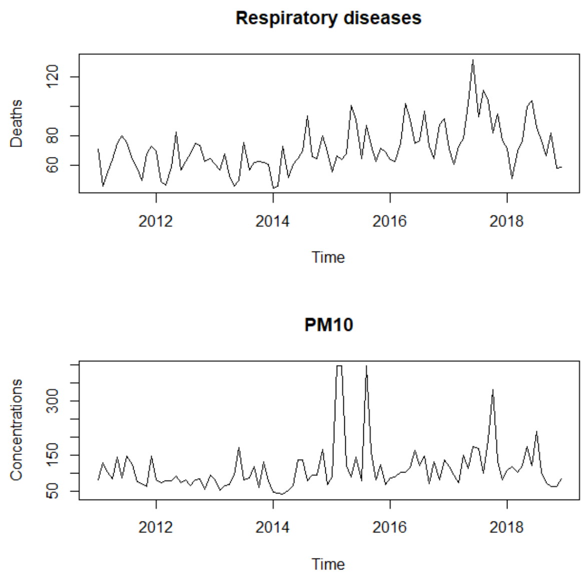
A real data analysis was realized to evaluate the impact of the Particulate Material (PM_{10}) on the monthly number of deaths by respiratory diseases between 2011 to 2018 ($n = 96$) in the Great Vitoria region (GVR), Brazil, which is a port and industrialized region, densely populated in the state of Espírito Santo, with approximately 1,900,000 inhabitants. Although the atmosphere is composed of many gases and particulate matter, only PM_{10} was considered because the data quality of other contaminants during the period was too poor. The PM_{10} are microscopic solid particles and liquid droplets suspended in the air, with a diameter of 10 micrometers (μm) 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)).

A significant correlation was observed between the number of deaths and the maximum monthly concentrations of PM_{10} in the atmosphere ($\rho = 0.45$). Imputation data were performed before fitting the model to handle the missing observations presented in the PM_{10} series. We used the multivariate imputation by chained equation method

(MICE), proposed by van Buuren and Oudshoorn (2000).

Figure 1 presents the series of deaths caused by respiratory diseases and concentrations of PM_{10}. The number of deaths shows a positive trend and seasonal behavior. Furthermore, the PM_{10} concentration also presents a positive trend and three peaks. These aberrant observations can be considered additive outliers.

Figure 1: Time series of the number of deaths by respiratory diseases and concentrations of PM_{10} in the metropolitan area of Vitoria, Brazil.



The modeling considered a slightly positive trend in the number of deaths, and to handle the annual seasonality, sine and cosine functions were incorporated. The model is written as

$$\eta_t = \beta_1 x_{t,1} + \beta_2 \text{trend} + \beta_3 \sin(2\pi t/12) + \beta_4 \cos(2\pi t/12) + Z_t, \tag{25}$$

where t is the month number, $x_{t,1}$ is the PM_{10} concentrations and Z_t is the autocorrelation structure of the GLARMA model.

The classic GLARMA Poisson (Davis et al. (2003)) and the robust approach proposed in Section 3.1 were adjusted, and their parameter estimation was compared. Table 9 presents the estimates $\hat{\beta}_i$'s of the parameters β_i's in the classic model. All the estimates were significant at the 5% level of significance, except β_1, the coefficient related to the PM_{10} levels in the atmosphere.

Table 10 presents the estimates of the robust GLARMA model. All the estimates were significant at the 5% level of significance.

Figure 2 plots the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) of the residuals in the classic and robust GLARMA models.

Table 9: Parameter estimates of the classic GLARMA model fitted to the number of deaths caused by respiratory diseases.

	Intercept	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
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

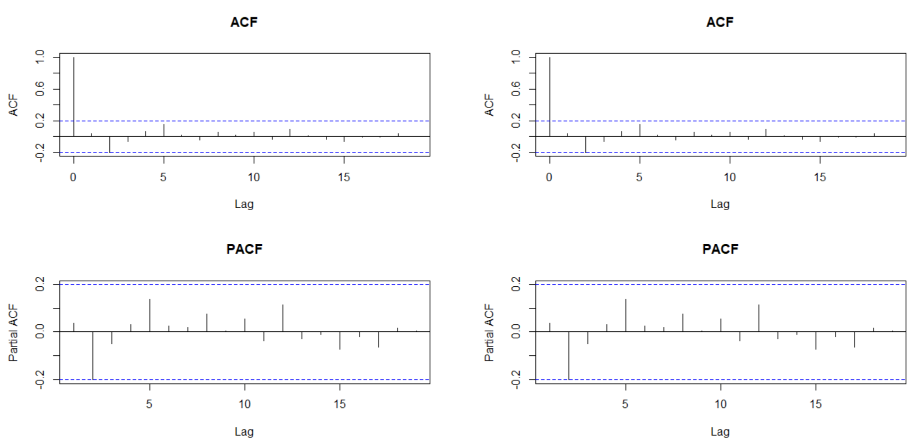
Table 10: Parameter estimates of the robust GLARMA model fitted to the number of deaths by respiratory diseases.

	Intercept	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
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

These plots show no difference with white noise, indicating a reasonable adjustment in both approaches.

Figure 2: Sample ACF and PACF of the residuals in the classic and robust GLARMA models.

(a) Classic GLARMA model. (b) Robust GLARMA model.



The parameter estimation in Tables 9 and 10 shows that although there is a significant correlation between PM_{10} concentrations and the monthly number of deaths in the period, in the classic GLARMA model, the parameter related to the pollutant was not significant at 5% level of significance. However, in the robust approach, the impact of the PM_{10} was significant, which means that this pollutant contributes significantly to the increase in deaths caused by respiratory diseases. It is essential to observe that the value of parameter β_1 seems to be underestimated in the classic model ($\hat{\beta}_{1(classic)} = 0.0003$), once the robust estimate was three times this value ($\hat{\beta}_{1(robust)} = 0.0009$).

To evaluate the performance of the GLARMA robust proposal in this real data analysis, we removed the peaks in PM_{10} concentration measurements and replaced them with its mean value. Then the classic GLARMA model was adjusted to the series without the extreme values. The robust approach is expected to behave approximately the same as this model. The values of the parameter estimates in Table 11 are similar to those in Table 10, with emphasis on β_1, the main affected by the presence of outliers, confirming our expectation.

Table 11: Parameter estimates of the classic GLARMA model fitted to the number of deaths by respiratory diseases.

	Intercept	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Estimate	4.0196	0.0008	0.0033	-0.0540	-0.0874
Standard error	0.0447	0.0003	0.0005	0.0220	0.0225
p-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 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 ζ-unit change in the X_i, keeping the other covariates fixed.

For Poisson regression, the RR is given by

$$\hat{\text{RR}}_{X_i}(\zeta) = \exp(\hat{\beta}_i \zeta) \tag{26}$$

The approximate confidence interval (CI) at an α significance level in the GLARMA with Poisson marginal distribution is

$$\widehat{\text{CI}}\{RR_{x_i}(\zeta)\} = \exp\left\{\zeta\left(\widehat{\beta}_i - z_{\alpha/2}\,\text{se}(\widehat{\beta}_i); \widehat{\beta}_i + z_{\alpha/2}\,\text{se}(\widehat{\beta}_i)\right)\right\}, \tag{27}$$

where $\widehat{\beta}_i$ is the conditional maximum likelihood estimator $\widehat{\beta}_{i,n}$ of β_i, se$(\widehat{\beta}_i)$ is the estimated standard deviation of $\widehat{\beta}_i$, and $z_{\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution.

The RR of the air pollutants is important information for the regulatory agencies to quantify the impact of these contaminants on the population's health. Table 12 presents the estimated RR and CIs for PM_{10}. The CIs were calculated with the bootstrap approach proposed by Camara et al. (2022) and the asymptotic approximation (equation (27)), with $\alpha = 5\%$. The $\hat{\text{RR}}$ was significant considering the classic and robust approaches (the value one does not belong to the intervals). In addition, the asymptotic and bootstrap CIs were equivalent, indicating that the classic model underestimated the relative risk. This result is in agreement with that observed in the simulation study.

Table 12: Estimated RR and 95% CI for PM_{10} in the classic and robust GLARMA models.

PM_{10}	Classic GLARMA	Robust GLARMA
$\widehat{RR}$	1.0187	1.0497
$\widehat{\text{CI}}$ asymptotic	[1.0001;1.0376]	[1.0305;1.0692]
$\widehat{\text{CI}}$ bootstrap	[1.0004;1.0379]	[1.0217;1.0794]

6 Conclusions

This work proposed a robust approach for the GLARMA model, introduced by Davis et al. (2003). This methodology is based on the robustification of the quasi-likelihood function using M-estimator to control deviations on response variable and weight functions to limit leverage points on covariates.

The simulation study showed that additive outliers could widely affect the classic GLARMA. The robust proposal behaves approximately like the classical approach in the absence of outliers. At the same time, for contaminated data, the parameter estimation was almost as good as the classic method applied to 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_{10} in the populations' health. This analysis showed that the RR is underestimated by the classic method, which means ignoring the impact of more than 100% of the exposure on the outcome. The numerical study agrees with this observation. The RR observed indicated that the PM_{10} contributed significantly to the increase of deaths by respiratory disease in the region.

Appendix 1

The derivative of constant $a(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ is

$$a'(\boldsymbol{\delta}) = \frac{\partial a(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = \frac{1}{n}\sum_{t=1}^{n}\left[(\mathrm{A})(\mathrm{B}) + (\mathrm{C})(\mathrm{D})\right], \tag{28}$$

where

(A)

$$c\left\{\sum_{k=j_2+1}^{\infty}\frac{1}{k!}\left[e^{-e^{W_t}}(e^{W_t})^k\frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}(k-e^{W_t}))\right] - \sum_{m=0}^{j_1}\frac{1}{m!}\left[e^{-e^{W_t}}(e^{W_t})^m\frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}(m-e^{W_t}))\right]\right\}$$
$$+\left(\frac{1}{2}(e^{W_t})^{1/2}\frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}\right)\left[\left(\frac{e^{-\mu_t}\mu_t^{j_1}}{j_1!}\right) - \left(\frac{e^{-\mu_t}\mu_t^{j_2}}{j_2!}\right)\right]$$
$$+(e^{W_t})^{1/2}\left\{\frac{1}{j_1!}\left[e^{-e^{W_t}}(e^{W_t})^{j_1}\frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}(j_1-e^{W_t})\right] - \frac{1}{j_2!}\left[e^{-e^{W_t}}(e^{W_t})^{j_2}\frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}(j_2-e^{W_t})\right]\right\}$$

(B)

$$(e^{W_t})^{1/2}\frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}$$

(C)

$$c\left[\sum_{t=j_2+1}^{\infty}\frac{e^{-\mu_t}\mu_t^k}{k!} - \sum_{m=0}^{j_1}\frac{e^{-\mu_t}\mu_t^m}{m!}\right] + (e^{W_t})^{1/2}\left[\frac{e^{-\mu_t}\mu_t^{j_1}}{j_1!} - \frac{e^{-\mu_t}\mu_t^{j_2}}{j_2!}\right]$$

(D)

$$\frac{1}{2}(e^{W_t})^{1/2}\frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}\frac{\partial W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^T} + (e^{W_t})^{1/2}\frac{\partial^2 W_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^T}$$

For $r_t \leq c$, the first derivative of $Q(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ is given by

$$\begin{aligned}\frac{\partial Q(\delta)}{\partial \delta} &= \sum_{t=1}^{n}\left[\left(\frac{Y_t-\mu_t}{\mu_t^{1/2}}\right)\frac{1}{\mu_t^{1/2}}w(\mathbf{X}_t)\mu_t\frac{\partial W_t(\boldsymbol{\delta})}{\partial \delta} - a(\delta)\right] \\ &= \sum_{t=1}^{n}\left[(Y_t-\mu_t)\,w(\mathbf{X}_t)\frac{\partial W_t(\boldsymbol{\delta})}{\partial \delta} - a(\delta)\right].\end{aligned}$$

Thus, the second derivative of $Q(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ is

$$\begin{aligned}\frac{\partial^2 Q(\delta)}{\partial\delta\partial\delta^T} &= \sum_{t=1}^{n}\left\{w(\mathbf{X}_t)\left[(Y_t-\mu_t)'\frac{\partial W_t(\boldsymbol{\delta})}{\partial\delta}+(Y_t-\mu_t)\frac{\partial^2 W_t(\boldsymbol{\delta})}{\partial\delta\delta^T}\right]-a'(\delta)\right\}\\ &= \sum_{t=1}^{n}\left\{w(\mathbf{X}_t)\left[-\mu_t\frac{\partial W_t(\boldsymbol{\delta})}{\partial\delta}\frac{\partial W_t(\boldsymbol{\delta})}{\partial\delta^T}+(Y_t-\mu_t)\frac{\partial^2 W_t(\boldsymbol{\delta})}{\partial\delta\delta^T}\right]-a'(\delta)\right\}\\ &= \sum_{t=1}^{n}\left\{(Y_t-\mu_t)\frac{\partial^2 W_t(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}^T}w(\mathbf{X}_t)-\mu_t\frac{\partial W_t(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}}\frac{\partial W_t(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}^T}w(\mathbf{X}_t)-a'(\boldsymbol{\delta})\right\}.\end{aligned}$$

For $r_t > c$, the first derivative of $Q(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ is given by

$$\begin{aligned}\frac{\partial Q(\delta)}{\partial\delta} &= \sum_{t=1}^{n}\left[c\text{sign}\left(\frac{Y_t-\mu_t}{\mu_t^{1/2}}\right)\frac{1}{\mu_t^{1/2}}w(\mathbf{X}_t)\mu_t\frac{\partial W_t(\boldsymbol{\delta})}{\partial\delta}-a(\delta)\right]\\ &= \sum_{t=1}^{n}\left[c\text{sign}\left(\frac{Y_t-\mu_t}{\mu_t^{1/2}}\right)\mu_t^{1/2}w(\mathbf{X}_t)\frac{\partial W_t(\boldsymbol{\delta})}{\partial\delta}-a(\delta)\right].\end{aligned}$$

Let $r_t = \frac{Y_t-\mu_t}{\mu_t^{1/2}}$, the $\frac{\partial^2 Q(\delta)}{\partial\delta\partial\delta^T}$ is given by

$$\sum_{t=1}^{n}\left(w(\mathbf{X}_t)\left\{\left[c\left(\frac{\partial}{\partial\delta}\text{sign}(r_t)\right)\left(\mu_t^{1/2}\frac{\partial W_t(\boldsymbol{\delta})}{\partial\delta}\right)\right]+c\text{sign}(r_t)\left[\frac{1}{2}\mu_t^{1/2}\frac{\partial W_t(\boldsymbol{\delta})}{\partial\delta}\frac{\partial W_t(\boldsymbol{\delta})}{\partial\delta^T}+\mu_t^{1/2}\frac{\partial^2 W_t(\boldsymbol{\delta})}{\partial\delta\partial\delta^T}\right]\right\}-a'(\delta)\right)$$

where

$$\frac{\partial}{\partial\delta}\text{sign}(r_t) = 2\delta(r_t),$$

and $\delta(.)$ is the Dirac delta function. By definition, $\delta(r_t) = 0$ if $r_t \neq 0$. As $r_t > c, c > 0$, $\delta(r_t) = 0$. Then

$$\frac{\partial^2 Q(\delta)}{\partial\delta\partial\delta^T} = \sum_{t=1}^{n}\left\{w(\mathbf{X}_t)c\text{sign}(r_t)\left[\frac{1}{2}\mu_t^{1/2}\frac{\partial W_t(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}}\frac{\partial W_t(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}^T}+\mu_t^{1/2}\frac{\partial^2 W_t(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}^T}\right]-a'(\boldsymbol{\delta})\right\}.$$

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 partially supported by CentraleSupélec and by the iCODE Institute, a research project of the IDEX Paris-Saclay, and by the Hadamard Mathematics LabEx (LMH) through the grant number ANR-11-LABX-0056-LMH in the Programme des Investissements d'Avenir.

References

D. Andrews. A robust method for multiple linear regression. *Technometrics*, 16(4): 523–531, 1974.

Z. Bai, C. Rao, and Y. Wu. M-estimation of multivariate linear 257 regression parameters under a convex discrepancy function. *Statistica Sinica*, 2(1), 1992.

J. Ballesteros-Cánovas, D. Trappmann, J. Madrigal-González, N. Eckert, and M. Stoffel. Climate warming enhances snow avalanche risk in the western himalayas. *Proceedings of the National Academy of Sciences (PNAS)*, 115(13), 2018.

A. Beaton and J. Tukey. The fitting of power series, meaning polynomials illustrated on band-spectroscopic data. *Technometrics*, 16(2):147–185, 1974.

M. Benjamin, R. Rigby, and D. Stasinopoulos. Generalized autoregressive moving average models. *J. Am. Stat. Assoc.*, 98(461):214–223, 2003.

C. Bergmeir and J. Benitez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 2012.

C. Bergmeir, R. Hyndman, and B. Koo. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics and Data Analysis*, 120:70–83, 2018.

G. Box and G. Jenkins. *Time series analysis: forecasting and control.* Holden-Day, San Fracisco, 1976.

O. Bustos and V. Yohai. Robust estimates for arma models. *Journal of American Statistical Association*, 81:155–168, 1986.

A. J. Camara, V. Reisen, G. Franco, and P. Bondon. Glarma model and bootstrap approaches: An application to respiratory diseases and air pollutants. *In submission process*, 2022.

E. Cantoni and E. Ronchetti. Robust inference for generalized linear models. *J. Am. Stat. Assoc.*, 96(455):1022–1030, 2001.

R. Carroll and A. Welsh. A note on asymmetry and robustness in linear regression. *American Statistician*, 42:285–287, 1986.

I. Chang, G. Tiao, and C. Chen. Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2):193–204, 1988.

C. Chen and L.-M. Liu. Joint estimation of model parameters and outlier effects in time series. *J. Am. Stat. Assoc.*, 88(421):284–297, 1993.

R. Chen, C. C., J. Tan, J. Cao, W. Song, X. Xu, C. Jiang, M. W., C. Yang, B. Chen, Y. Gui, and H. Kan. Ambient air pollution and hospital admission in shanghai, china. *J. Hazard. Mater.*, 181(1):234–240, 2010.

E. Chi. M-estimation in cross-over trials. *Biometrics*, 50(2):486–493, 1994.

D. Cox. Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, 8:93–115, 1981.

R. Davis, Y. Wang, and W. Dunsmuir. *Modelling Time Series of Count Data. In: S. Ghosh (ed) Asymptotics, Nonparametrics, and Time Series.* CRC Press, New York, 1999.

R. Davis, W. Dunsmuir, and S. Streett. Observation driven models for Poisson counts. *Biometrika*, 90(4):777–790, 2003.

R. Davis, W. Dunsmuir, and S. Streett. Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodology and Computing in Applied Probability*, 7(2):149–159, 2005.

R. Davis, K. Fokianos, S. Holan, and H. Joe. Count time series: A methodological review. *Journal of the American Statistical Association*, 116(535):1533–1547, 2021.

J. Dennis and R. Welsch. Techniques for nonlinear least squares and robust regression. *Communications in Statistics-simulation and Computation*, 7(4):345–459, 1978.

D. Dockery and C. Pope. *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, 1996.

W. Dunsmuir. *from: Handbook of discrete-valued time series, Handbook of Modern Statistical Methods*. CRC Press, London, 2015.

K. Fokianos and J. Tjosthein. Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102(3):563–578, 2011.

A. Fox. Outliers in time series. *Journal of the Royal Statistical Society: Series B*, 34(3): 350–363, 1972.

C. Froes, V. Camara, P. Landrigan, and L. Claudio. Systematic review of children's environmental health in brazil. *Annals of Global Health*, 82(1):132–148, 2016.

F. Hampel. Contributions to the theory of robust estimation. *Ph.D. thesis, University of California, Berkeley*, 1968.

F. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393, 1974.

F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley and Sons, Inc, 1986.

P. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101, 1964.

P. Huber. *Robust Statistics*. Wiley, New York, 1981.

S. Karami, M. Karami, G. Roshanaei, and H. Farsan. Association between increased air pollution and mortality from respiratory and cardiac diseases in tehran: Application of the glarma model. *Iranian Journal of Epidemiology*, 12(4):36–43, 2017.

J.-Y. Kim, H.-Y. Kim, D. Park, and Y. Chung. Modelling of fault in rpm using the glarma and ingarch model. *Eletronics Letters*, 54(5):297–299, 2018.

S. Kitromilidou and K. Fokianos. Robust estimation methods for a class of log-linear count time series models. *Journal of Statistical Computation and Simulation*, 86(4): 740–755, 2016.

H. Künsch. Infinitesimal robustness for autoregressive processes. *Annals of Statistics*, 12:843–863, 1984.

H. Künsch, L. Stefanski, and R. Carroll. Conditionally unbiased bounded-influence estimation in general regression models with applications to generalized linear models. *Journal of the American Statistical Association*, 84:460–466, 1989.

J. Ledolter. The effect of additive outliers on the forecast from arima models. *International Journal of Forecasting,*, 5:231–240, 1989.

T. Li. Laplace periodogram for time series analysis. *Journal of the American Statistical Association*, 103:757–768, 2008.

M. Lippmann. Toxicological and epidemiological studies of cardiovascular effects of ambient air fine particulate matter (pm2.5) and its chemical components: Coherence and public health implications. *Critical Reviews in Toxicology*, 44(4):299–347, 2014.

S. Lo and E. Ronchetti. Robust and accurate inference for generalized linear models. *Journal of Multivariate Analysis*, 100:2126–2136, 2009.

C. Mallows. On some topics in robustness. *Technical Memorandum, Murray Hill, N.J., Bell Telephone Laboratories*, 1975.

R. Maronna, R. Martin, and V. Yohai. *Robust Statistics: Theory and Methods*. John Wiley & Sons, Chichester, 2006.

R. Martin and V. Yohai. *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, 1985.

J. Nelder and R. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A*, 135(3):370–384, 1972.

B. Ostro, G. Eskeland, J. Sánchez, and T. Feyzioglu. Air pollution and health effects: A study of medical visits among children in santiago, chile. *Environmental Health Perspectives*, 107(1):69–73, 1999.

B. Ostro, L. Roth, B. Malig, and M. Marty. he effects of fine particle components on respiratory hospital admissions in children. *Environmental Health Perspectives*, 117 (3):475–480, 2009.

E. Peitzsch, G. Pederson, K. Birkeland, J. Hendrikx, and D. Fagre. Climate drivers of large magnitude snow avalanche years in the u.s. northern rocky mountains. *Scientific Reports*, 11(10032), 2021.

C. Pope and D. Dockery. Health effects of fine particulate air pollution: lines that connect. *Journal of the Air Waste Management Association*, 23(2):307–318, 2018.

C. Pope, M. Thun, M. Namboodiri, D. Dockery, J. Evans, F. Speizer, and C. Heath. Particulate air pollution as a predictor of mortality in a prospective study of u.s. adults. *American Journal Respiratory Critical Care Medicine*, 151:669–674, 1995.

V. Reisen, C. Levy-Leduc, and M. Taqqu. An m-estimator for the long-memory parameter. *Journal of Statistical Planning and Inference*, 187:44–55, 2017.

P. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 79(388):871–880, 1984.

P. Rousseeuw. *Multivariate Estimation With High Breakdown Point. In: W. Grossmann, G. Pflug, I. Vincze and W. Wertz (eds) Mathematical Statistics and Applications.* Reidel Publishing Company, Dordrecht, 1985.

A. Ruckstuhl and A. Welsh. Robust fitting of the binomial model. *CMA, Australian National University, manuscript*, 1999.

T. Rydberg and N. Shephard. Dynamics of trade-by-trade price movements:decomposition and models. *Journal of Financial Econometrics*, 1(1):2–25, 2003.

A. Sarnaglia, V. Reisen, C. Levy-Leduc, and P. Bondon. M-regression spectral estimator for periodic arma models. an empirical investigation. *Stochastic Environmental Research and Risk Assessment*, 35:653–664, 2021.

J. Schwartz. Harvesting and long-term exposure effects in the relationship between air pollution and mortality. *Am. J. Epidemiol.*, 151(5):440–448, 2000.

D. Simpson, D. Ruppert, and R. Carroll. On one-step gm estimates and stability of inferences in linear regression. *On One-Step GM Estimates and Stability of Inferences in Linear Regression*, 87(418):439–450, 1992.

J. Souza, V. Reisen, G. Franco, M. Ispány, P. Bondon, and J. Santos. Generalized additive models with principal component analysis: an application to time series of respiratory disease and air pollution data. *Journal of the Royal Statistical Society: Series C*, 67 (2):453–480, 2018.

J. Streett, R. Carroll, and D. Ruppert. A note on computing robust regression estimates via iterative reweighted least squares. *Journal of the American Statistical Association*, 42(2):152–154, 1988.

J. Tukey. *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, 1960.

M. Valdora and V. Yohai. Robust estimators for generalized linear models. *Journal of Statistical Planning and Inference*, 146:31–48, 2014.

S. van Buuren and K. Oudshoorn. Multivariate imputation by chained equations:mice v1.0 user's manual. *TNO Prevention and Health*, 2000.

W. Wu. M-estimation of linear models with dependent errors. *The Annals of Statistics*, 35(2):495–521, 2007.

S. L. Zeger and B. Qaqish. Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, 44(4):1019–1031, 1988.