mathematics

MDPI

Article

Combining Generalized Linear Autoregressive Moving Average and Bootstrap Models for Analyzing Time Series of Respiratory Diseases and Air Pollutants

Ana Julia Alves Camara [1], Valdério Anselmo Reisen [2,*], Glaura Conceicao Franco [3] and Pascal Bondon [4]

[1] Department of Statistics, Federal University of Espirito Santo, Av. Fernando Ferrari, 514, Vitoria 29075-910, Brazil; ana.j.camara@ufes.br
[2] PPGEA (Graduate Program in Environmental Engineering), Federal University of Espirito Santo, Av. Fernando Ferrari, 514, Vitoria 29075-910, Brazil
[3] Department of Statistics, Federal University of Minas Gerais, Av. Antonio Carlos 6627, Belo Horizonte 31270-901, Brazil; glaura@est.ufmg.br
[4] Laboratoire des Signaux et Systèmes, CentraleSupélec, CNRS, Université Paris-Saclay, 3 Rue Joliot Curie, 91192 Gif-sur-Yvette, France; pascal.bondon@centralesupelec.fr
* Correspondence: valderio.reisen@ufes.br

Abstract: The generalized linear autoregressive moving-average model (GLARMA) has been used in epidemiology to evaluate the impact of pollutants on health. These effects are quantified through the relative risk (RR) measure, which inference can be based on the asymptotic properties of the maximum likelihood estimator. However, for small series, this can be troublesome. This work studies different types of bootstrap confidence intervals (CIs) for the RR. The simulation study revealed that the model parameter related to the data's autocorrelation could influence the intervals' coverage. Problems could arise when covariates present an autocorrelation structure. To solve this, using the vector autoregressive (VAR) filter in the covariates is suggested.

Keywords: time series of counts; INAR models; integer-valued data; respiratory diseases; air pollution

MSC: 62M10

Academic Editors: Luiz Koodi Hotta and Pedro Luiz Valls Pereira

Received: 10 December 2024
Revised: 28 February 2025
Accepted: 2 March 2025
Published: 5 March 2025

Citation: Camara, A.J.A.; Reisen, V.A.; Franco, G.C; Bondon, P. Combining Generalized Linear Autoregressive Moving Average and Bootstrap Models for Analyzing Time Series of Respiratory Diseases and Air Pollutants. *Mathematics* **2025**, *13*, 859. https://doi.org/10.3390/math13050859

1. Introduction

Time series of counts are non-Gaussian processes composed of non-negative integers. These series can be found in different scientific areas, such as economics, medicine, agriculture, social and physical sciences, and sports. Some examples are the daily number of hospital admissions for a disease, the number of car accidents in a region, and the number of transactions of a given stock observed in one minute. In recent decades, various approaches have emerged to model correlated count series. Recently, ref. [1] presented a review of these methodologies and discussed recent developments on this topic.

The impact of air pollution on population health has been the subject of many studies in the last decades; see [2–5], among others. The massive and continuous development of cities and communities leads to urbanization and industrialization. However, it can cause environmental and health problems as many activities generate residues that affect inhabitants' quality of life [6].

Epidemiological studies have consistently provided evidence of the association between daily levels of pollutant concentration and hospital admissions, morbidity, and mortality, mainly caused by respiratory and cardiovascular diseases; see [7,8] for references.

Epidemiological data are frequently treated as time series of counts because they record the relative frequency of certain events in successive time intervals. In this context, many authors have been using the generalized additive model (GAM) [9] with Poisson marginal distribution to quantify the association between the effects of air pollution on health.

Alternative methods such as wavelet transforms have shown promise in capturing complex temporal patterns in time series, particularly those with non-stationary behavior or cyclical components, such as air pollution data. Ref. [10] demonstrated the application of wavelet transforms combined with a weighting forecasting model to predict air pollution levels, highlighting their robustness in handling complex environmental data sets. Furthermore, last-squares wavelet analysis (LSWA), as described in [11], offers a robust framework for time-frequency analysis, even with high noise levels and missing data. LSWA has been successfully applied to simulated time series for component estimation at various confidence intervals and forecasting purposes, demonstrating its versatility and reliability.

Despite widespread use, care is required when applying the GAM to time series. The GAM assumes that errors are mutually independent and, therefore, cannot capture the time-dependent structure of the observations. One way to circumvent this is to use the generalized linear autoregressive moving-average (GLARMA) model, proposed by [12], which adds an ARMA structure to generalized linear models (GLMs) [13]. The GLM is an extension of the Gaussian linear model, where the distribution of the response variable belongs to the exponential family.

The GLARMA model allows for modeling correlated observations from the exponential family. Despite its complexity in estimating general models, this methodology has been widely applied in various fields (see, e.g., [14–19], among others). As discussed by [1], the GLARMA family is one of the most flexible and easily adaptable count models, effectively balancing parameter-driven and observation-driven approaches. Besides the GLARMA model, different approaches to modeling count time series were also suggested in the literature. Ref. [20] proposed a quasi-likelihood approach to time series regression, which was generalized by [21] and called generalized autoregressive moving-average models (GARMA).

Ref. [22] proposed the Autoregressive Conditional Poisson model (ACP), able to model overdispersion, and ref. [23] proposed log-linear models for time series. Recently, ref. [24] proposed a model using the GAM with autoregressive moving-average terms (GAM-ARMA). This procedure can model the temporal correlation structure and estimate nonlinear associations between the covariates and the response variable. Following the Bayesian approach, ref. [25] used state-space models with conjugate prior distributions, where the counts are modeled as a Poisson distribution, and ref. [26] proposed a family of non-Gaussian state-space models.

In epidemiology, the relative risk (RR) is a standard statistical measure used to quantify the relationship between contaminant levels and adverse health effects. Beyond point estimation, epidemiologists are also interested in interval estimation for RR. Although the asymptotic properties of the maximum likelihood estimators are not yet established for the general case in GLARMA models, confidence intervals (CIs) for relative risks are often computed under normality assumptions. While this approach is not problematic for large sample sizes, it may be unreliable for small samples.

Still underexplored in studies related to air quality, bootstrap CIs can be calculated without any assumptions about data distribution. This procedure, introduced by [27] initially for independent observations, has been extended to more general situations. Refs. [28–30] used a classic model-based approach, where bootstrap samples are generated using the estimated model structure and independently and identically distributed (i.i.d.) sampling residues. Ref. [31] used the classical idea of the residual-based

bootstrap for semiparametric generalized models. Ref. [32] proposed a nonparametric bootstrap for the stationary process, the blockwise bootstrap, where blocks of consecutive observations are resampled. This method is robust against misspecification models, but the block dependence is neglected. Ref. [33] proposed another nonparametric methodology, the sieve bootstrap, an approach to real-valued time series where an autoregressive process of order p, AR(p), is adjusted to the observations to approximate the real data distribution. In this procedure, the bootstrap samples are obtained by resampling from the centered residuals of the autoregressive process fitted. Due to the popularity of the sieve bootstrap and its simple implementation, in recent years, many authors have proposed methodologies to adapt it to time series of counts. The most simple and probably obvious approximation is to centralize the counts and adjust the AR(p) process to the centralized observations. However, this approach leads to valid bootstrap approximation only for a limited number of cases (for details, see [34]).

In the last decades, the most prominent works proposing approaches for the sieve bootstrap considered the integer-valued autoregressive (INAR) time series, motivated by the fact that the AR and INAR processes share the same autocorrelation function. In the INAR model, the dependent count observation, $\{Y_t\}_{t\in\mathbb{Z}}$, is expressed as a function of the p preceding values plus an innovation $\{\epsilon_t\}_{t\in\mathbb{Z}}$, with $\epsilon_t \sim G$, where G is a distribution with range $\mathbb{N}_0$. Refs. [35–37] proposed relevant alternatives to constructing appropriate confidence intervals' bootstrap for the INAR models. Ref. [34] proposed an efficient method for bootstrapping INAR models. This paper addressed a critical literature review, highlighting the bootstrapping count time series challenge and the limitations of the proposed techniques considering INAR models. Ref. [34] introduced a general INAR-type procedure, where the INAR(p) model is adjusted to the count time series and a marginal distribution $\hat{G}$ is attributed to the bootstrap innovations. The authors proved the consistency of this procedure, assuming parametric and nonparametric distributions for the bootstrap innovations. In addition, they investigated the performance of this procedure by analyzing the coverage of 95% CIs for diverse statistics based on the observations. Their numerical simulation illustrates the superiority of the proposed method over the blockwise, sieve, and Markov bootstraps.

Based on the above discussion and the real data problem, this paper compares and evaluates different bootstrap approaches for computing confidence intervals for relative risk using GLARMA models, focusing on their coverage rates. For this, three of those techniques cited previously were examined: the classic model-based approach, based on the work of [31], the well-known sieve bootstrap, and the INAR-type bootstrap. Confidence intervals assuming Gaussian distribution for the estimators were also calculated.

Although the authors of [34] have already verified the superiority of their methodology over the sieve bootstrap for statistics based on the dependent count observations, here we aim to compare the performance of the investigated methods applied to the GLARMA model. An extensive simulation study was performed considering distinct sample sizes, types of covariates, and different autocorrelation structures for a response variable autocorrelated and conditioned to the past, following a Poisson distribution. The objective was to verify if any change in the characteristics of the model's covariates or the complexity of the autocorrelation structure in the GLARMA model could impact the coverage or amplitude of the calculated intervals.

A real-time series was analyzed to illustrate the findings of the numerical simulations. We computed confidence intervals for the RR to quantify the impact of air pollutants on the number of chronic obstructive pulmonary disease cases in the metropolitan area of Belo Horizonte, Brazil. This study addresses a critical gap in the literature, as few works have explored bootstrap methods with GLARMA models in the context of relative risk

estimation. Moreover, the motivation stems from the practical importance of accurately quantifying uncertainty in real-world time series data, such as those related to public health or environmental studies, where reliable interval estimates are crucial for decision making.

This paper is organized as follows. Section 2 presents the GLARMA model and some essential properties regarding the model and the parameter estimation. Section 3 discusses the bootstrap for time series of counts, where three approaches are introduced. Simulation studies are performed in Section 4, considering different scenarios and sample sizes. A real data analysis is carried out in Section 3.2, and Section 4 presents the final considerations of the work.

2. Methodology

2.1. The Generalized Linear Autoregressive Moving-Average Model

The GLARMA models are a class of observation-driven state-space models. The state process consists of linear regression and observation-driven components comprising an autoregressive moving-average filter of past predictive residuals.

Let $\{Y_t\} := \{Y_t\}_{t\in\mathbb{Z}}$ be the observations and $\mathcal{F}_{t-1} = (Y^{(t-1)}, X^{(t)})$, where $Y^{(t-1)}$ is the past of the counting process and $X^{(t)}$ is the past and present of the regressor variables. Conditional on $\mathcal{F}_{t-1}$, the observations are independent and have a distribution in the exponential family with density

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

where $b(\cdot)$ and $c(\cdot)$ are known real functions and $\{W_t\}$ summarizes the information in $\mathcal{F}_{t-1}$; see [1,38].

Ref. [12] considered that the specification of $\{W_t\}$ in (1) is given by

$$W_t = \mathbf{X}_t^T \boldsymbol{\beta} + \sum_{i=1}^{\infty} \gamma_i e_{t-i}, \tag{2}$$

where $\mathbf{X}_t = (1, X_{1,t}, X_{2,t}, \ldots, X_{k,t})$ is the vector of covariates of dimension $k+1$, $\boldsymbol{\beta}$ is a $(k+1)\times 1$ vector of unknown coefficients to be estimated, and the standard residuals e_t are defined as

$$e_t = \frac{(Y_t - e^{W_t})}{e^{\lambda W_t}}, \tag{3}$$

for $\lambda \in (0, 1]$.

The infinite moving-average weights γ_i in (2) can be specified in terms of an autoregressive moving-average (ARMA) filter:

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

where the autoregressive and moving-average components $\boldsymbol{\phi}(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$ and $\boldsymbol{\theta}(B) = (1 + \theta_1 B + \ldots + \theta_q B^q)$ are polynomials with roots outside the unit circle, $\boldsymbol{\gamma}$ is the parameter vector formed by ϕs and θs, and B is the backshift operator of the form $B^k(Z_t) = Z_{t-k}$. Defining $W_t = \mathbf{X}_t^T \boldsymbol{\beta} + Z_t$, where $Z_t = \sum_{i=1}^{\infty} \gamma_i e_{t-i}$, for $t \leq 0$ and $e_t = 0, Z_t = 0$ and for $t > 0$, the process Z_t is computed according to the following ARMA-like recursions:

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

Remark 1. *Let $e_s = 0$ and $Y_s = 0$ for $s \leq 0$. Under these conditions, it is easy to show that the e_t form a martingale difference sequence with zero mean, variance given by $E(\mu_t^{1-2\lambda}), t \geq 1$,*

and $Cov(e_t, e_s) = 0$ *for* $t \neq s$. *In addition, for any* $\lambda \in (0, 1]$, $E(W_t) = x_t^T\beta, Var(W_t) = \sum_{i=1}^{\infty} \gamma_i^2 E(\mu_{t-i}^{1-2\lambda})$, *and, for* $s = t + h, h > 0$, $Cov(W_t, W_{t+h}) = \sum_{i=1}^{\infty} \gamma_i \gamma_{i+h} E(\mu_{t-i}^{1-2\lambda})$. *For more details, see Section 2.2 in [12].*

Remark 2. *For* $\lambda = 1$, *ref. [12] showed that considering the simplest model, where* $Y_t|\mathcal{F}_{t-1} \sim Poi(\mu_t)$ *and* $W_t = \beta_0 + \gamma(Y_{t-1} - e^{W_{t-1}})e^{-\lambda W_{t-1}}$, *assuming* $p = 0$ *and* $q = 1$, *the process* $\{W_t\}$ *has a unique stationary distribution and is uniformly ergodic. For* $\frac{1}{2} \leq \lambda \leq 1$, $\{W_t\}$ *is bounded in probability and therefore has a stationary distribution, yet the uniqueness of this distribution is currently unknown.*

Define $\delta = (\beta_0, \beta_1, \ldots, \beta_k, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^T$ as the parameter vector of the model (2). Let $y_1, y_2, \ldots, y_n$ be a sample from the process $\{Y_t\}$; the log-likelihood of $\{Y_t\}$ conditional to $\mathcal{F}_{t-1}$ is given by

$$L(\delta) = \sum_{t=1}^{n} \log f(y_t|W_t(\delta)). \tag{6}$$

For the exponential family, the log-likelihood becomes

$$L(\delta) = \sum_{t=1}^{n} \{y_t W_t(\delta) - a_t b(W_t(\delta)) + c_t\}.$$

As a particular case, for $Y_t|\mathcal{F}_{t-1} \sim \text{Poi}(\mu_t)$, where $\mu_t = e^{W_t}$, the log-likelihood is given by

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

The log-likelihood function can be maximized using Newton–Raphson iteration and Fisher scoring approximation procedures. Defining the first and second derivatives of the log-likelihood by $d(\delta) = \partial L(\delta)/\partial\delta$ and $D_{NR}(\delta) = \partial^2 L(\delta)/\partial\delta\partial\delta^T$, we have

$$d(\delta) = \sum_{t=1}^{n} \big(y_t - a_t \dot{b}(W_t)\big) \frac{\partial W_t}{\partial \delta} \tag{8}$$

and

$$D_{NR}(\delta) = \sum_{t=1}^{n} \big(y_t - a_t \dot{b}(W_t)\big) \frac{\partial^2 W_t}{\partial\delta\partial\delta^T} - \sum_{t=1}^{n} a_t \ddot{b}(W_t) \frac{\partial W_t}{\partial \delta} \frac{\partial W_t}{\partial \delta^T}. \tag{9}$$

At the true parameter of δ, $E(y_t - a_t \dot{b}(W_t))|\mathcal{F}_{t-1}) = 0$, the expected value of the first summation in (9) is zero, which motivates the Fisher scoring approximation:

$$D_{FS}(\delta) = -\sum_{t=1}^{n} a_t \ddot{b}(W_t) \frac{\partial W_t}{\partial \delta} \frac{\partial W_t}{\partial \delta^T}. \tag{10}$$

Note that $E(D_{NR}(\delta)) = E(D_{FS}(\delta))$; however, these expectations cannot be calculated in closed form. Thus, the Newton–Raphson (using D_{NR}) and the Fisher scoring (using D_{FS}) methods are used to maximize the log-likelihood function.

Ref. [1] discussed that the consistency and asymptotic properties of the maximum likelihood estimator $\hat{\delta}$ were proven only for the simplest model cited in Remark 2. For this special case, stationarity and ergodicity were established rigorously. In general, for inference purposes, it is assumed that the central limit theorem holds.

$$\hat{\delta} \approx N(\delta, \mathbf{\Omega}),$$

where the covariance matrix of the estimators $\mathbf{\Omega}$ is given by $-[D_{NR}(\delta)]^{-1}$ using the Newton–Raphson iteration and $-[D_{FS}(\delta)]^{-1}$ using the Fisher scoring approximation. The GLARMA

package [39] provides functions for estimation based on the generalized linear autoregressive moving-average class for discrete-valued time series with regression variables.

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. The authors of [40] present its mathematical representation:

$$\frac{E(Y|X_i = \zeta, X_j = x_j, i \neq j)}{E(Y|X_i = 0, X_j = x_j, i \neq j)}.$$

For Poisson regression, the RR is given by

$$\mathrm{RR}_{X_i}(\zeta) = \exp(\beta_i \zeta) \tag{11}$$

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

$$\widehat{\mathrm{RR}}_{X_i}(\zeta) = \exp\left\{\zeta\left(\widehat{\beta}_i - z_{\alpha/2}\,\mathrm{se}(\widehat{\beta}_i); \widehat{\beta}_i + z_{\alpha/2}\,\mathrm{se}(\widehat{\beta}_i)\right)\right\}, \tag{12}$$

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

2.2. Bootstrap for Count Time Series

Initially proposed by [27] for independent variables, the bootstrap is a resampling method that attributes measures of accuracy to statistical estimates. It is a computer-based procedure that approximates the theoretical distribution by the empirical distribution of a finite sample of observations.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote the random sample and its observed realizations, respectively, from a distribution F. A bootstrap sample $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ is obtained by randomly sampling n times, with replacement, from the original data points $x_1, x_2, \ldots, x_n$. That is, the bootstrap sample is drawn from the empirical distribution F^*. For any statistic u computed from the original sample data, it is possible to define a statistic u^* by the same formula but calculate it using the resampled data.

Ignoring the temporal correlation of the time series can impact the estimations. In this context, many approaches have been proposed, standing out from the classic model-based [28–30], the blockwise [32], and the sieve [33] bootstraps. In the residual model-based approach, the bootstrap replications are based on i.i.d. resampling of residuals. The blockwise and sieve bootstraps are nonparametric procedures that are robust against model misspecification, as discussed in [33]. In the first case, blocks of consecutive observations are resampled. In the sieve bootstrap, the autoregressive process of order p, AR(p), is adjusted to the observations to approximate the actual data distribution.

However, the challenge arises when the bootstrap is applied to non-Gaussian responses. Several authors have studied the bootstrap involving time series of counts in recent decades. Ref. [31] proposed a methodology derived from the classic model based on semiparametric generalized models. Refs. [34–37] proposed approaches to adapt the sieve bootstrap considering integer-valued autoregressive (INAR) time series.

In this section, three bootstrap procedures for count time series are discussed: the residual model-based approach of [31], an adaptation of the sieve bootstrap for count observations, and the proposal of [34] for the INAR(p) process.

2.2.1. Classic Model-Based Bootstrap

This procedure is used in regression problems, in which it is assumed that the model is correctly specified and the error terms in the model are independent and identically distributed. The basic idea of the bootstrap with parametric assumptions is to obtain residuals using the estimated parametric model and then generate bootstrap samples using the estimated model structure and i.i.d. resampling of residuals. Assume that Y_t, given the past history, $\mathcal{F}_{t-1}$, has a distribution in the exponential family with $\mu_t = g\{X_t^T\beta + Z_t\}$, $t \in \mathbb{Z}$, where g is a known link function and the component Z_t is defined in Equation (5). The bootstrap procedure works as follows:

(1) Compute the residuals $\hat{e}_t = (Y_t - \hat{\mu}_t)/\hat{\mu}_t^{\lambda}$.
(2) Resample the estimated residuals with replacements generating the bootstrap residual samples $(e_1^*, \ldots, e_n^*)$.
(3) Generate bootstrap observations $Y_1^*, \ldots, Y_n^*$ according to

$$Y_t^* = \hat{\mu}_t + e_t^* \hat{\mu}_t^{\lambda}.$$

In Algorithm 1, we present the procedure for calculating the parametric bootstrap for regression models, referred to here as the classic model-based bootstrap:

Algorithm 1 Parametric Bootstrap Procedure for Regression Models

Input: Observed data $\{Y_t, X_t\}_{t=1}^n$, link function g, number of bootstrap samples $n_{\text{bootstrap}}$
Output: Bootstrap samples $\{Y_t^*\}_{t=1}^n$
Estimate the model parameters $\hat{\beta}$ and compute $\hat{\mu}_t = g(X_t^T\hat{\beta} + Z_t)$ for $t = 1, \ldots, n$
Compute the residuals:
    $\hat{e}_t = \frac{Y_t - \hat{\mu}_t}{\hat{\mu}_t^{\lambda}}, \quad t = 1, \ldots, n$
For each bootstrap sample $b = 1, 2, \ldots, n_{\text{bootstrap}}$:
for $t = 1$ to n **do**
    Resample residuals e_t^* with replacements from $\{\hat{e}_t\}_{t=1}^n$
    Generate bootstrap observation:
        $Y_t^* = \hat{\mu}_t + e_t^* \hat{\mu}_t^{\lambda}$
end for
Store the bootstrap sample $Y^* = \{Y_t^*\}_{t=1}^n$
end for
return Bootstrap samples $\{Y_t^*\}_{t=1}^n$

This procedure is model-based, and misspecification problems can be observed, e.g., biased parameters and inconsistent standard errors. The sieve bootstrap, presented below, follows the same strategy of fitting a parametric model and resampling the residuals but approximating an infinite-dimensional nonparametric model by a sequence of finite-dimensional parametric models.

2.2.2. Sieve Bootstrap

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a real-valued stationary process, and denote $Y_1, Y_2, \ldots, Y_n$ a sample from this process. The basic idea of the sieve bootstrap [33] is to adjust an autoregressive process of order p to the data:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + \xi_t, \quad t \in \mathbb{Z}, \tag{13}$$

with increasing order p as the sample size n increases. The estimated coefficients $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_n$ are then used to compute the residuals $\hat{\xi}_t$. These residuals are centered,

$\tilde{\xi}_t = \hat{\xi}_t - (n-p)\sum_{t=p+1}^{n} \hat{\xi}_t$, and the bootstrap sample $Y_1^*, Y_2^*, \ldots, Y_n^*$ is constructed according to the recursion

$$Y_t^* = \hat{\alpha}_1 Y_{t-1}^* + \hat{\alpha}_2 Y_{t-2}^* + \ldots + \hat{\alpha}_p Y_{t-p}^* + \tilde{\xi}_t^*, \quad t \in \mathbb{Z}, \tag{14}$$

where $\tilde{\xi}_t^*$ is a random sample with replacement of the centered residuals.

The sieve bootstrap was initially proposed for real-valued data, but a simple approximation can be performed by ignoring the discrete nature of the process. Ref. [34] cited this approach as follows:

(1) Compute the centered observations $X_t = Y_t - \bar{Y}$, where $\bar{Y} = 1/n \sum_{t=1}^{n} Y_t$.
(2) Fit an AR(p) to the centered data $X_1, \ldots, X_n$

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \xi_t,$$

where the estimated AR coefficients $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$ can be obtained, e.g., from Yule–Walker estimates.
(3) Compute the estimated residuals $\hat{\xi}_t = X_t - \sum_{i=1}^{n} \hat{\alpha}_i X_{t-i}$, and center them obtaining $\tilde{\xi}_t$ as presented previously.
(4) Generate bootstrap observations $X_1^*, \ldots, X_n^*$ according to

$$X_t^* = \hat{\alpha}_1 X_{t-1}^* + \hat{\alpha}_2 X_{t-2}^* + \ldots + \hat{\alpha}_p X_{t-p}^* + \tilde{\xi}_t^*,$$

where $\tilde{\xi}_t^*$ randomly and uniformly resamples from the centered residuals.
(5) The AR bootstrap sample of the original process $\{Y_t\}_{t\in\mathbb{Z}}$ can be calculated as $Y_t^* = X_t^* + \bar{Y}$.

Ref. [41] showed that the sieve bootstrap captures the autocovariance structure of a process and will always be consistent for the sample mean and any statistic that depends exclusively on the autocovariance structure of the process under mild conditions. Algorithm 2 presents the procedure for calculating the sieve bootstrap:

2.2.3. INAR-Type Bootstrap

Ref. [34] proposed the procedure based on the INAR model introduced by [42,43] and extended by [44,45].

The integer-valued autoregressive process of order p, denoted by INAR(p), expresses the value of the variable of interest at time t as a function of the p preceding values and an innovation in the following way:

$$Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + \ldots + \alpha_p \circ Y_{t-p} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{15}$$

where $\{Y_t\}_{t\in\mathbb{Z}}$ is a sequence of non-negative integer-valued variables, ϵ_t is an i.i.d non-negative integer-valued random variable with finite mean μ_ϵ and variance σ_ϵ^2, and $\alpha = (\alpha_1, \ldots, \alpha_p)^T \in (0,1)^p$, such that $\sum_{i=1}^{p} \alpha_i < 1$. The operator "$\circ$" in (15) is called the *binomial thinning operator* and $\alpha_i \circ Y_{t-i} \sim \text{Bin}(Y_{t-i}, \alpha_i)$, where $\text{Bin}(n, \pi)$ denotes the binomial distribution with parameters n and π. Due to the random thinning operator, the INAR models are nonlinear, contrary to the sieve bootstrap, which belongs to the linear time series process class. It is important to observe that the INAR innovations ϵ_t and the sieve errors e_t were distinguished here. As these procedures share the same autocorrelation function, the INAR coefficients can be estimated using techniques from classical time series analysis such as Yule–Walker, Least-Squares, or Maximum-Likelihood estimators. The authors of [45] proved that if $\sigma_\epsilon^2 = \text{Var}(\epsilon_t) < \infty$ the autocorrelation of an INAR(p) process corresponds to that of an autoregressive process of order p, AR(p). However, unlike the sieve bootstrap, the procedure proposed by [34] does not explicitly use the residuals from

the fitted model. Let Y_t be a non-negative integer-valued time series. The general INAR bootstrap scheme is defined as follows:

(1) Fit an INAR(p) process as in Equation (15) to obtain the estimates of $\alpha_1, \ldots, \alpha_p$ for the INAR coefficients;
(2) Specify the marginal distribution $\hat{G}$ for the innovations ϵ_t;
(3) Generate the bootstrap samples according to

$$Y_t^* = \hat{\alpha}_1 \circ^* Y_{t-1}^* + \ldots + \hat{\alpha}_p \circ^* Y_{t-p}^* + \epsilon_t^*, \tag{16}$$

where $\circ^*$ denotes the mutually independent bootstrap binomial thinning operations and (ϵ_t^*) are i.i.d. random variables following the distribution $\hat{G}$.

Algorithm 2 Sieve Bootstrap for Time Series Data ([33])

Input: Original time series data $Y = \{Y_1, Y_2, \ldots, Y_n\}$,
  p: Order of the autoregressive model,
  B: Number of bootstrap samples,
  k: Number of sieve bootstrap repetitions.
Output: $\{Y_b^*\}_{b=1}^B$: Bootstrap samples of the time series.
for $b = 1$ to B **do**
  Compute the mean of the original data $\bar{Y} = \frac{1}{n}\sum_{t=1}^n Y_t$
  Center the observations: $X_t = Y_t - \bar{Y}$, for $t = 1, 2, \ldots, n$
  Fit an autoregressive model of order p (AR(p)) to the centered data:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \xi_t, \quad t = p+1, \ldots, n$$

  where ξ_t are the residuals.
  Estimate the AR coefficients $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$ using methods like Yule–Walker
  Compute the residuals: $\hat{\xi}_t = X_t - \hat{\alpha}_1 X_{t-1} - \hat{\alpha}_2 X_{t-2} - \cdots - \hat{\alpha}_p X_{t-p}$
  Center the residuals: $\tilde{\xi}_t = \hat{\xi}_t - \frac{1}{n}\sum_{t=1}^n \hat{\xi}_t$
  for $k = 1$ to k **do**
    Resample the centered residuals $\tilde{\xi}_t^*$ by random sampling with replacements from $\{\tilde{\xi}_1, \ldots, \tilde{\xi}_n\}$
    Generate the bootstrap series:

$$X_t^* = \hat{\alpha}_1 X_{t-1}^* + \hat{\alpha}_2 X_{t-2}^* + \cdots + \hat{\alpha}_p X_{t-p}^* + \tilde{\xi}_t^*$$

  for $t = p+1, \ldots, n$
  end for
  Reconstruct the original scale: $Y_t^* = X_t^* + \bar{Y}$
  Store the bootstrap sample: $Y_b^* = \{Y_1^*, Y_2^*, \ldots, Y_n^*\}$
end for
return Bootstrap samples $\{Y_b^*\}_{b=1}^B$

The estimation of the parameters $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$ and the choice of the distribution $\hat{G}$, in steps 1 and 2, respectively, can be calculated following parametric and semiparametric approaches. Here, the parametric perspective was considered. In this case, under some assumptions related to the marginal distribution G of the innovation process (see [34]), the bootstrap innovations $\epsilon_1^*, \ldots, \epsilon_n^*$ can be easily generated from $G_{\hat{\theta}}$ following the steps above. In Algorithm 3, we present the procedure for calculating the bootstrap INAR(1):

Algorithm 3 Bootstrap INAR(1)—([34])

Input: Observed time series $\{y_t\}_{t=1}^{n}$, order of the process $p = 1$, number of bootstrap samples $n_{\text{bootstrap}}$
Output: Bootstrap samples $\{Y_t^*\}_{t=1}^{n}$
Estimate the INAR(1) coefficients $\hat{\alpha}_1$ using the observed data
    $\hat{\alpha}_1 = \text{Estimate}(\{y_t\})$
Specify the marginal distribution $\hat{G}$ for the innovations ϵ_t
    For example, assume $\epsilon_t \sim \text{Poisson}(\mu_\epsilon)$, where $\mu_\epsilon = \frac{1}{n}\sum_{t=1}^{n} y_t$
For each bootstrap sample $b = 1, 2, \ldots, n_{\text{bootstrap}}$:
for t = 1 to n **do**
    Initialize Y_t^* for $t = 1$;
        $Y_1^* = y_1$ (initial condition)
    for t = 2 to n **do**
        Generate bootstrap innovation $\epsilon_t^* \sim \hat{G}$
        Apply the binomial thinning operation:
            $Y_t^* = \text{Binomial}(Y_{t-1}^*, \hat{\alpha}_1) + \epsilon_t^*$
    end for
    Store the bootstrap sample $Y^* = \{Y_t^*\}_{t=1}^{n}$
end for
return Bootstrap samples $\{Y_t^*\}_{t=1}^{n}$

2.2.4. Bootstrap Confidence Intervals

Bootstrap confidence intervals can be computed using simple percentiles, bias-corrected percentile limits, bias-corrected and accelerated percentiles (BCa), and the Student's t method, among other proposals. Here, we will focus on the approach used by [34] in their simulations. To compute the bootstrap confidence interval for a parameter θ, these authors first calculated a centering measure cent($\hat{\theta}^*$) and the centered bootstrap estimates $\hat{\theta}^*_{cent} := \hat{\theta}^* - \text{cent}(\hat{\theta}^*)$. Then, the bootstrap confidence interval was calculated using the $(1-\alpha/2)$- and $\alpha/2$-quantiles from $\hat{\theta}^*_{cent}$ as follows:

$$\left[\hat{\theta} - q_{1-\alpha/2}\left(\hat{\theta}^*_{cent}\right); \hat{\theta} - q_{\alpha/2}\left(\hat{\theta}^*_{cent}\right)\right], \tag{17}$$

where $\hat{\theta}$ is the estimate obtained from the sample.

As the main objective of this paper is to calculate bootstrap confidence intervals for the RR of the variable $X_i, i = 1, \ldots, k$, the CIs were constructed as follows:

$$\widehat{\text{RR}}_{X_i}(\zeta) = \exp\left\{\zeta\left(\hat{\beta}_i - q_{1-\alpha/2}\left(\hat{\beta}_i{}^*_{cent}\right); \hat{\beta}_i - q_{\alpha/2}\left(\hat{\beta}_i{}^*_{cent}\right)\right)\right\}, \tag{18}$$

where $\hat{\beta}_i$ is the i-th estimated coefficient, $\hat{\beta}_i{}^*$ is the bootstrap estimate, and $\hat{\beta}_i{}^*_{cent} = \left(\hat{\beta}_i{}^* - \overline{\hat{\beta}_i{}^*}\right)$, with $\overline{\hat{\beta}_i{}^*}$ being the mean of $\hat{\beta}_i{}^*$. The interquartile variation of X_i is given by ζ, and α is the significance level.

In Algorithm 4, we present the procedure for calculating the bootstrap confidence intervals for the RR:

Algorithm 4 Bootstrap Confidence Interval for the Relative Risk (RR)

Input: Sample estimates $\hat{\beta}_i$, bootstrap estimates $\hat{\beta}_i^*$, significance level α, and interquartile variation ζ
Output: Bootstrap confidence intervals for the relative risk $\widehat{\mathrm{RR}}_{X_i}(\zeta)$
Compute the centering measure for the bootstrap estimates:
    $\mathrm{cent}(\hat{\beta}_i^*) = \frac{1}{n_{\text{bootstrap}}} \sum_{b=1}^{n_{\text{bootstrap}}} \hat{\beta}_i^*$
Compute the centered bootstrap estimates:
    $\hat{\beta}_{i\,\text{cent}}^* = \hat{\beta}_i^* - \mathrm{cent}(\hat{\beta}_i^*)$
For each bootstrap sample $b = 1, 2, \ldots, n_{\text{bootstrap}}$:
for $i = 1$ to k **do**
    Compute the confidence intervals for the relative risk $\widehat{\mathrm{RR}}_{X_i}(\zeta)$:
        $\widehat{\mathrm{RR}}_{X_i}(\zeta) = \exp\{\zeta(\hat{\beta}_i - q_{1-\alpha/2}(\hat{\beta}_{i\,\text{cent}}^*); \hat{\beta}_i - q_{\alpha/2}(\hat{\beta}_{i\,\text{cent}}^*))\}$
end for
return Bootstrap confidence intervals $\widehat{\mathrm{RR}}_{X_i}(\zeta)$

3. Results

3.1. Simulation Study

A simulation study was conducted to evaluate and compare the performance of the bootstrap approaches presented in Section 2.2. As many studies use asymptotic confidence intervals for the RR based on the Gaussian distribution, this interval was also considered for comparison purposes. This analysis focused on the confidence intervals for the relative risk calculated from the GLARMA(1,0) Poisson model, given by

$$Y_t|\mathcal{F}_{t-1} \sim \text{Poisson}(\mu_t)$$
$$\ln(\mu_t) = \beta_0 + \beta X_t + Z_t,$$

where Z_t is given by

$$Z_t = \phi[Z_{t-1} + (Y_{t-1} - \mu)\mu^{-\lambda}]. \tag{19}$$

In this simulation study, we analyze time series data representing pollutant levels, such as carbon monoxide (CO), as input variables. The pollutant values, denoted as X_t, are combined with fixed parameters (β and ϕ) to generate the response variable Y_t, which reflects health outcomes. We model this relationship using a GLARMA(1,0) Poisson model.

In the data-generating process, the initial value for Y_0 was set as the mean of the response variable Y, while Z_0 was set to zero. Regarding the burn-in period, in our implementation, we discarded the first 30% of the generated iterations to ensure that the sampled values more accurately reflected the stationary properties of the process.

The output data consist of bootstrap samples of the estimated parameters generated using the procedures outlined in Section 2. These bootstrap samples are subsequently used to compute confidence intervals for the relative risk (RR). Figure 1 provides a general illustration of this procedure.

The effectiveness of the bootstrapping method for non-Gaussian distributions depends on the sample size, as the central limit theorem ensures a better approximation of the sampling distribution as the sample size increases ([46], p. 153). Although no universally defined minimum sample size exists, previous studies suggest that samples with at least 30 to 50 observations are often sufficient for reasonable estimates.

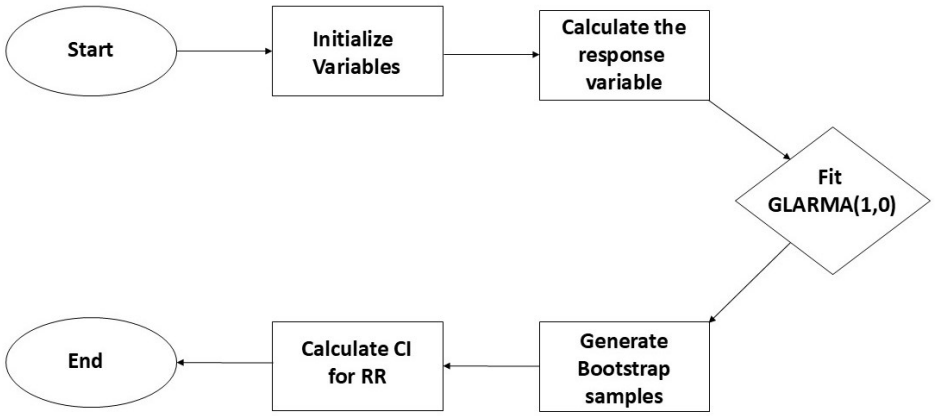

Figure 1. Flowchart of simulation study.

3.1.1. Large Samples

For this simulation study, three scenarios were considered:

- S1: $\beta = (\beta_0, \beta_1)$ and $X_t = (1, X_{1,t})^T$, where $X_{1,t} = t/n$.
- S2: $\beta = (\beta_0, \beta_1)$ and $X_t = (1, X_{1,t})^T$, where $X_{1,t}$ is an independent random vector in time and $X_{1,t} \sim N(0, 1)$.
- S3: $\beta = (\beta_0, \beta_1)$ and $X_t = (1, X_{1,t})^T$, where $X_{1,t} \sim \text{AR}(1)$, with the autoregressive parameter assuming values 0.2, 0.5, and 0.8.

These scenarios were selected with specific objectives: Scenario 1 uses the covariate from [12] for comparison purposes. Scenario 2 considers a covariate with no temporal dependence, allowing us to examine the impact of the absence of a temporal structure on the bootstrap intervals. Scenario 3 incorporates a covariate with distinct levels of temporal dependence, aiming to assess the influence of more complex temporal structures on the variables. Of the three scenarios, Scenario 3 is the most closely related to real-world situations, as pollutants typically exhibit temporal dependence.

The considered sample size n was equal to 1000. The values of ϕ in Equation (19) were fixed at 0.2, 0.4, and 0.6, and the nominal level for the confidence intervals was fixed at 95%. The parameter λ assumed the value 1.0. Although [12] showed that, for $\lambda = 1$, only in the simplest model, the process $\{W_t\}$ has a stationary and ergodic distribution, the numerical simulations revealed that, even for complex models, this value of λ provides better estimates. For all scenarios, the parameter values used in the simulations were chosen for simplicity, with $\beta_0 = \beta_1 = 1.0$. Additional simulations conducted with different values of β produced similar results, indicating that the findings are not sensitive to the specific choice of parameter values.

The classic model-based bootstrap considers the three steps presented in Section 2.2.1 to construct the confidence intervals. For the sieve and INAR(1) cases, the steps shown in Sections 2.2.2 and 2.2.3 were applied to the count time series Y_t, and then the GLARMA Poisson model was fitted considering the bootstrap samples as the response variables. In addition, for the INAR(1) bootstrap, an INAR process having Poisson-distributed innovations was assumed with $\epsilon_t \sim \text{Poisson}(\lambda)$ and $\hat{\lambda} = \bar{Y}(1 - \hat{\alpha})$, where $\hat{\alpha}$ was obtained by Yule–Walker estimation.

The Monte Carlo simulations were repeated 500 times with 500 bootstrap replications. The asymptotic confidence interval was estimated as in Equation (12). All the codes were written in the R language and are available from the authors upon request.

- Scenario 1

Table 1(a) presents the mean and standard deviation of the 500 Monte Carlo estimates of parameters β_0, β_1, and ϕ in scenario 1 (S1). For β_0 and β_1 parameters, the mean of

the estimates was close to the real values, especially when $\phi = 0.2$ or 0.4. Conversely, the standard deviation shows a consistent increase with the value of ϕ. Parameter ϕ is better estimated for small values. Table 1(b) presents the 95% confidence intervals for the RR of the covariate $X_{1,t}$. The values in square brackets are the calculated intervals' average lower and upper limits. The results indicate that classic and sieve bootstrap methods exhibit a decline in coverage as ϕ increases. Specifically, for the classic bootstrap, the coverage drops from 0.884 to 0.729, while for the sieve bootstrap, the decrease is even more pronounced, from 0.927 to 0.570. Notably, for $\phi = 0.2$, the sieve bootstrap still achieves a coverage rate close to the nominal level of 0.95, suggesting that it may be a reasonable choice in this scenario. However, for higher values of the autoregressive parameter, both methods present considerably lower coverage, indicating potential limitations in their performance under strong dependence. In contrast, the INAR(1) bootstrap and the asymptotic confidence intervals maintained a coverage rate of approximately 95% for all values of ϕ.

Scenario 1 studied the impact of the same covariate considered in [12], although the authors only evaluated cases where the time correlation is a moving-average process of order 1. Real data sets commonly present an autoregressive autocorrelation structure. In this case, S1 showed that even when the time correlation structure is complex (e.g., $\phi = 0.6$), for deterministic covariates ($X_{1,t} = t/n$), the asymptotic theory and the INAR(1) bootstrap presented coverage rates close to the nominal level of the confidence intervals.

Table 1. (**a**) The mean and standard deviation of the Monte Carlo estimates of parameters β_0, β_1, and ϕ in scenario 1 (S1). (**b**) The coverage and the average lower and upper limits of the 95% confidence intervals for the RR = $\exp(\zeta\beta_1)$ (S1).

(**a**)						
	ϕ **= 0.2**		ϕ **= 0.4**		ϕ **= 0.6**	
	Mean	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**
β_0	0.992	0.042	0.990	0.052	0.932	0.068
β_1	1.011	0.066	1.012	0.082	1.078	0.102
ϕ	0.200	0.030	0.392	0.026	0.553	0.102
(**b**)						
RR = 1.28	$\phi = 0.2$		$\phi = 0.4$		$\phi = 0.6$	
Classic bootstrap	0.884 [1.218; 1.370]		0.802 [1.226; 1.385]		0.729 [1.222; 1.401]	
Sieve bootstrap	0.927 [1.217; 1.368]		0.786 [1.227; 1.378]		0.570 [1.265; 1.419]	
INAR(1) bootstrap	0.958 [1.196; 1.386]		0.944 [1.193; 1.427]		0.930 [1.161; 1.435]	
Asymptotic	0.962 [1.199; 1.385]		0.948 [1.181; 1.428]		0.934 [1.154; 1.484]	

- Scenario 2

In the parameter estimation presented in Table 2(a), the mean of the estimates was close to the real values of all parameters, except when $\phi = 0.6$. It can also be seen that the standard deviations were much affected by the increase in ϕ. Table 2(b) presents the 95% confidence intervals for the RR in scenario 2 regarding the covariate $X_{1,t}$. Table 2(b) shows that for the classic bootstrap, the coverage rate was close to 1 for all values of ϕ, which means that almost 100% of the intervals contain the true relative risk value. Regarding the sieve bootstrap, for $\phi = 0.2$ and 0.4, the coverage rate was also close to 1. Meanwhile, for $\phi = 0.6$, the coverage rate decreased to 0.849. The INAR(1) bootstrap and the asymptotic intervals had similar performance, with coverages close to 0.95 for $\phi = 0.2$ and 0.4 and considerably below the nominal level for 0.6. It should be pointed out that the INAR(1) bootstrap always presented coverages closer to 0.95 than the asymptotic interval, even for the $\phi = 0.6$ case.

Scenarios 1 and 2 showed that the coverage of the INAR bootstrap and asymptotic approaches were close to the nominal level for $\phi = 0.2$ and 0.4, where β_1 was appropriately estimated. In Tables 1(a) and 2(a), the mean of this parameter is close to the true values, and although the standard deviation goes up in Table 2(a), the coverage in scenario 2 is not impacted. However, for $\phi = 0.6$, the β_1 estimates were terrible, mainly in S2, and as the RR depends on β_1, the interval coverage was also impacted. Finally, it is essential to observe that the coverage intervals are unsuitable for classic and sieve bootstraps even when the β_1 estimates are reasonable.

Table 2. (**a**) The mean and standard deviation of the Monte Carlo estimates of parameters β_0, β_1, and ϕ in scenario 2 (S2). (**b**) The coverage and the average lower and upper limits of the 95% confidence intervals for the RR = $\exp(\zeta\beta_1)$ (S2).

(**a**)						
	ϕ **= 0.2**		ϕ **= 0.4**		ϕ **= 0.6**	
	Mean	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**
β_0	0.995	0.026	1.010	0.285	1.201	1.046
β_1	1.002	0.015	0.992	0.139	0.951	0.299
ϕ	0.199	0.017	0.380	0.077	0.430	0.192
(**b**)						
RR = 1.94	ϕ **= 0.2**		ϕ **= 0.4**		ϕ **= 0.6**	
Classic bootstrap	0.990 [1.827; 2.096]		1.000 [1.802; 2.086]		0.992 [1.778; 2.137]	
Sieve bootstrap	0.986 [1.834; 2.088]		0.989 [1.821; 2.093]		0.849 [1.908; 2.067]	
INAR(1) bootstrap	0.953 [1.863; 2.038]		0.948 [1.875; 2.051]		0.732 [1.989; 2.155]	
Asymptotic	0.968 [1.860; 2.041]		0.940 [1.887; 2.049]		0.638 [2.005; 2.143]	

- Scenario 3

In epidemiology, it is common for air pollutants to present temporal correlation. To simulate this behavior, in scenario 3, the covariate $X_{1,t}$ followed an autoregressive process of order 1:

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

where $X_{1,t}$ is an AR(1) process with autoregressive parameter φ and Z_t is defined by Equation (5). To evaluate the time structure's impact, the covariate's autoregressive parameter (φ) assumed values equal 0.2, 0.5, and 0.8. Table 3(a) presents the parameter estimates for all values of φ. For $\varphi = 2$, when focusing on β_1, the mean estimate of this parameter is close to the true value for $\phi = 0.2$, while the estimate becomes less accurate for $\phi = 0.4$ and $\phi = 0.6$, accompanied by an increase in the standard deviation.

In the case of $\varphi = 0.5$, shown in Table 3(a), the mean of the β_1 estimate remains close to the true value for $\phi = 0.2$. However, compared to the results for the exact value of ϕ in Table 3(a), there is a noticeable increase in the standard deviation. For both $\phi = 0.4$ and $\phi = 0.6$, the β_1 estimates deteriorate, and the standard deviation increases as the value of ϕ rises.

Table 3(a) also shows the parameter estimates when the covariate's autoregressive parameter (X_t) is $\varphi = 0.8$. Even for the smallest value of ϕ considered, the mean estimate of β_1 is poor, and the standard deviation is significantly high. For $\phi = 0.4$ and $\phi = 0.6$, the means of the β_1 estimates become much worse, and the standard deviation increases even further.

Table 3(b) presents the 95% confidence intervals (CIs) for the relative risk (RR) of the covariate $X_{1,t}$. For $\varphi = 0.2$, the coverage rate for the classic model-based bootstrap was

close to 1 for all values of ϕ. The performance of the asymptotic approach, sieve bootstrap, and INAR(1) bootstrap was similar, with the coverage decreasing for $\phi = 0.4$. For $\phi = 0.6$, all methods exhibited poor performance.

For $\varphi = 0.5$, as shown in Table 3(b), a similar performance was observed for the classic and sieve bootstraps, with the coverage rate equal to 100% for $\phi = 0.2$, followed by a drop in coverage as ϕ increased. Both methods also produced large confidence intervals. The coverage rate of the INAR(1) and asymptotic approaches was approximately 95% for $\phi = 0.2$. For $\phi = 0.4$, these rates decreased, with the INAR(1) bootstrap maintaining the highest coverage. Again, for $\phi = 0.6$, all methods showed substantial deviations from the nominal coverage level.

For $\varphi = 0.8$, Table 3(b) shows that for $\phi = 0.2$, the classic model-based and sieve bootstraps had coverage rates close to 1, while the INAR(1) bootstrap and asymptotic approach exhibited coverage rates of 0.93 and 0.895, respectively. For $\phi = 0.4$ and $\phi = 0.6$, all methods saw a significant decline in the coverage rate, with the CIs from the asymptotic approach and INAR(1) bootstrap being the most affected.

Table 3. (**a**) The mean and standard deviation of the Monte Carlo estimates of parameters β_0, β_1, and ϕ in scenario 3 (S3). (**b**) The coverage and the average lower and upper limits of the 95% confidence intervals for the RR = $\exp(\zeta\beta_1)$ (S3).

(**a**)

		ϕ = 0.2		ϕ = 0.4		ϕ = 0.6	
		Mean	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**
$\varphi = 0.2$	β_0	0.978	0.038	0.811	0.275	1.246	0.795
	β_1	1.020	0.031	1.134	0.147	0.907	0.300
	ϕ	0.138	0.021	0.234	0.072	0.355	0.235
$\varphi = 0.5$	β_0	1.028	0.178	1.101	0.326	1.970	2.195
	β_1	0.974	0.159	0.936	0.216	0.657	0.696
	ϕ	0.190	0.041	0.306	0.157	0.148	0.167
$\varphi = 0.8$	β_0	1.127	1.038	2.048	1.509	6.593	3.359
	β_1	0.962	0.257	0.448	0.734	−0.478	0.514
	ϕ	0.175	0.067	0.052	0.098	0.001	0.008

(**b**)

RR = 4.03		$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$
$\varphi = 0.2$	Classic bootstrap	1.000 [3.428; 4.312]	1.000 [3.419; 4.528]	0.998 [3.386; 4.440]
	Sieve bootstrap	0.974 [3.960; 5.152]	0.892 [3.815; 5.052]	0.378 [3.529; 4.389]
	INAR(1) bootstrap	0.957 [3.850; 4.171]	0.921 [3.877; 4.167]	0.272 [3.439; 3.671]
	Asymptotic	0.956 [3.894; 4.193]	0.908 [3.910; 4.169]	0.182 [3.475; 3.634]
$\varphi = 0.5$	Classic bootstrap	1.000 [4.109; 5.149]	0.977 [4.078; 6.545]	0.802 [1.695; 6.159]
	Sieve bootstrap	1.000 [4.067; 5.718]	0.941 [3.934; 4.844]	0.382 [3.771; 4.571]
	INAR(1) bootstrap	0.921 [4.583; 4.938]	0.767 [4.371; 4.708]	0.145 [4.248; 4.554]
	Asymptotic	0.944 [4.574; 4.947]	0.751 [4.369; 4.711]	0.111 [4.316; 4.550]
$\varphi = 0.8$	Classic bootstrap	1.000 [6.043; 11.948]	0.791 [3.379; 13.769]	0.227 [0.626; 5.914]
	Sieve bootstrap	0.998 [6.899; 9.078]	0.426 [4.657; 5.989]	0.206 [2.730; 3.145]
	INAR(1) bootstrap	0.927 [8.132; 8.721]	0.176 [5.444; 5.773]	0.000 [1.170; 1.212]
	Asymptotic	0.895 [8.162; 8.688]	0.154 [5.276; 5.560]	0.000 [1.172; 1.211]

In general, we observed that the INAR method appears to perform better, alongside the asymptotic method, as they present narrower confidence intervals and coverage rates closer to the nominal value. On the other hand, the classic and sieve methods showed very poor coverage in some cases, with coverage rates close to 1, while the nominal value

was 0.95. This suggests that the INAR and asymptotic approaches are more reliable for estimating the relative risk in scenario 3, although the coverage rate was in general inferior than that observed in scenario 2.

The comparison between scenarios 2 and 3 indicates that time correlation in the covariates can impact the coverage rate of the confidence intervals; as the autoregressive structure becomes more complex, the interval coverage becomes smaller. Table 3(a) showed that the values of ϕ strongly impact the parameter estimation, and this effect becomes worse as this autoregressive parameter increases in the direction of the nonstationarity region, either in the covariate or in the Z_t component. It is easy to verify that for any $\lambda \in (0, 1]$, $\text{Var}(W_t) = \sum_{i=1}^{\infty} \gamma^2 \text{E}(\mu_{t-i}^{1-2\lambda})$, where the covariate $X_{1,t}$ is an independent random vector in time. However, for $X_{1,t} \sim \text{AR}(1)$, the variability of the state process $\{W_t\}$ increases, which inflates the model estimates, directly impacting the coverage rates of the RR.

Beyond the empirical investigations discussed here, scenarios with more complex model structures, such as bivariate time series, were also considered. As expected, the coverage rate was unsatisfactory. Therefore, the authors recommend applying the procedure proposed by [7] before implementing the bootstrap approaches discussed in this study in practical situations where covariates are time series. This is further explored in Section 3.2, within the real data analysis, where the covariates follow a vector of time series data.

3.1.2. Small Samples and ARMA Covariate

This work also studied another scenario considering small samples for the GLARMA(1,0) Poisson model. In this case, the sample size was equal to 50, and the single covariate $X_{1,t}$ was an ARMA(p, q) process:

$$Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\mu_t)$$
$$\ln(\mu_t) = \beta_0 + \beta_1 X_{1,t} + Z_t,$$

where $X_{1,t}$ is an ARMA process with autoregressive and moving-average parameters φ and θ and Z_t is defined by Equation (19). Three different ARMA processes were considered. They were chosen due to their temporal structure, which is similar to some atmospheric pollutants in real data:

- (1) $X_{1,t} \sim \text{ARMA}(1, 1)$, where $\varphi_1 = 0.8$ and $\theta_1 = 0.2$.
- (2) $X_{1,t} \sim \text{ARMA}(1, 1)$, where $\varphi_1 = 0.8$ and $\theta_1 = 0.4$.
- (3) $X_{1,t} \sim \text{ARMA}(2, 1)$, where $\varphi_1 = 0.5$, $\varphi_2 = 0.3$, and $\theta_1 = 0.4$.

Table 4 presents the 95% confidence intervals for the RR, where $X_{1,t} \sim \text{ARMA}(p, q)$, where $p = 1, 2$ and $q = 1$. Here, only the INAR(1) bootstrap and the asymptotic approach are compared, as these methodologies presented similar results in the previous simulation studies. The autoregressive parameter ϕ was fixed at 0.2 once the simulations presented the best adjustments at this value. In all cases, the INAR(1) bootstrap presented a coverage rate approximately equal to 0.95, while for the asymptotic approach, this rate was close to 0.90.

Table 4. The coverage rate of the 95% confidence intervals for the RR (n = 50).

	ARMA(1,1) $\boldsymbol{\varphi_1 = 0.8}$ **and** $\boldsymbol{\theta_1 = 0.2}$	**ARMA(1,1)** $\boldsymbol{\varphi_1 = 0.8}$ **and** $\boldsymbol{\theta_1 = 0.4}$	**ARMA(2,1)** $\boldsymbol{\varphi_1 = 0.5, \varphi = 0.3}$ **and** $\boldsymbol{\theta_1 = 0.4}$
INAR(1) bootstrap	0.958	0.970	0.942
Asymptotic	0.896	0.906	0.910

The INAR(1) bootstrap presented better results related to the coverage rate for the RR than the asymptotic approach, considering the ARMA process as a covariate. This analysis

is relevant in real data sets. The sample size is generally insignificant, and the air pollutants present complex structures, e.g., time correlation, high volatility, and peaks. In addition, the coverage rates were close to 0.95 in the INAR(1) bootstrap even for high values of the autoregressive parameter ($\varphi = 0.8$), different from those observed in the simulations of scenario 3 (S3), in Section 4.1, where the covariate was an AR(1) process.

3.2. Real Data Analysis

A real data analysis was proposed to study the impact of air pollutants on the monthly number of chronic obstructive lung disease (COPD) cases in Belo Horizonte, Brazil, between 2007 and 2013 ($n = 84$). Figure 2 presents the pollutants considered: Particulate Matter (PM_{10}), Nitrogen Monoxide (NO), Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), and Ozone (O_3). This figure shows an increasing trend in COPD cases, along with seasonality that may be semiannual and/or annual. The air pollutants show a generally constant trend, with occasional aberrant values. Ozone (O_3) appears to exhibit a more pronounced seasonality.

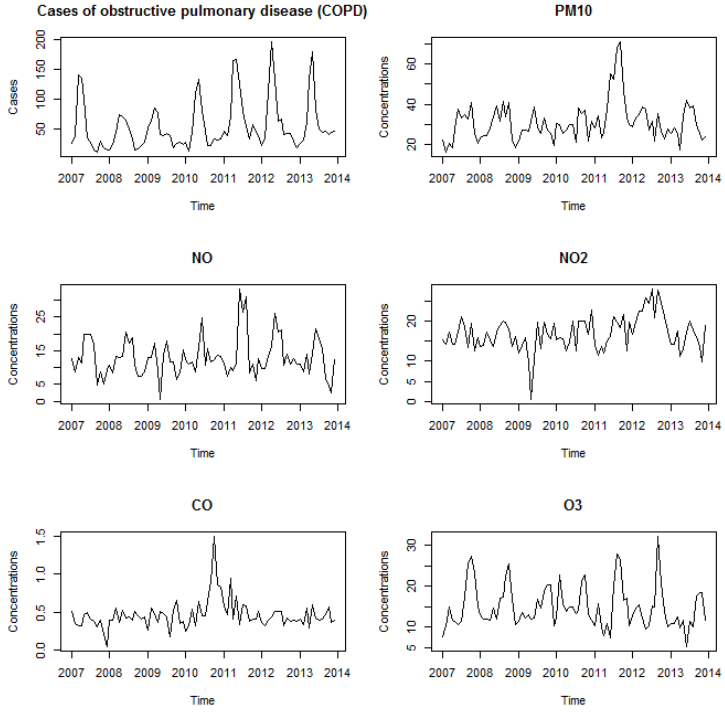

Figure 2. The time series of the number of COPD cases and concentrations of air pollutants in the metropolitan area of Belo Horizonte, Brazil.

Many covariates can lead to identification problems, and their correlation may imply multicollinearity. A significant correlation among some air pollutants was observed, while O_3 and NO presented the highest correlation with the response variable (COPD). Table 5 presents the correlation matrix among air pollutants and the number of chronic obstructive pulmonary cases. Principal component analysis (PCA) is a possible solution to this problem. This methodology explains a random vector's variance and covariance structure through linear combinations of the original variables ([47]). These combinations, called principal components (PCs), are not correlated with each other. The PCA methodology requires independent observations; however, according to [48], if the covariates present a time correlation, then the PCs are also autocorrelated. Based on this, the works of [7,8] proposed a hybrid model called GAM-PCA-VAR, where the time dependency of data is removed through the VAR process. The PCs are derived from the residuals of VAR, and the GAM model is adjusted with PCs as explanatory variables.

Table 5. Pearson correlation among pollutants and chronic obstructive pulmonary disease cases.

	COPD	**CO**	**PM_{10}**	**NO**	**NO_2**	**O_3**
COPD	1.00					
CO	0.03	1.00				
PM_{10}	0.08	0.15	1.00			
NO	0.30	0.10	0.47	1.00		
NO_2	−0.03	0.15	0.29	0.52	1.00	
O_3	−0.35	0.07	0.37	−0.19	0.24	1.00

This study estimated the impact of air pollutants on the occurrence of COPD using the previously mentioned procedure. To model this relationship, it is essential to apply methods that account for the time dependence inherent to COPD cases. The GLARMA model is an attractive option due to its flexibility and easy fit using statistical packages (see [39]). The time correlation structure of the contaminants was removed by applying the VAR filter, and the PCs were derived from the residuals of VAR. The GLARMA model was fitted using the PCs as covariates. All the principal components were considered, although the first three correspond to almost 80% of the entire structure of variability. The GLARMA Poisson model was adjusted as follows:

$$\begin{aligned}\ln(\mu_t) = \beta_1 * \mathrm{PC1}_t + \beta_2 * \mathrm{PC2}_t + \beta_3 * \mathrm{PC3}_t + \beta_4 * \mathrm{PC4}_t + \beta_5 * \mathrm{PC5}_t \\ + \beta_6 * \sin 6t + \beta_7 * \cos 6t + \beta_8 * \sin 12t + \beta_9 * \cos 12t \\ + \beta_{10} * \mathrm{trend} + Z_t,\end{aligned}$$

where

$$Z_t = \phi_1(Z_{t-1} + e_{t-1}) + \phi_2(Z_{t-2} + e_{t-2}).$$

The annual and semiannual seasonality in the response variable was incorporated into the model with sine and cosine functions. The modeling also included the trend present in the data. This component was modeled using a linear function of time, where a variable ranging from 1 to n was included to account for potential long-term patterns in the data.

Table 6 presents the estimates of the adjusted model, with the corresponding standard errors. We considered different autocorrelation structures and used the AIC and BIC measures to compare them. The best fit was obtained regarding the autoregressive parameter of order 2. All coefficients were significant at the 5% level of significance.

Confidence intervals for the relative risk of air pollutants were calculated under the assumption of normality (asymptotic) and using the INAR(1) bootstrap. The RR was computed based on the combination of the PCs. For more details about this estimation, see [7], Section 2.2.4. Algorithm 5 outlines the procedure for computing bootstrap confidence intervals.

In the application, we utilized a bootstrap procedure with 1000 iterations to ensure reliable and stable estimates, incorporating a burn-in period set to 30% of the total iterations. This was performed to allow the model to reach a stable state, particularly considering the strong temporal dependence in the data. By discarding the first 300 iterations, we minimized the potential influence of initial conditions on the parameter estimates, ensuring that the results reflected the true underlying process. This approach is especially crucial when the observed values of the series are significantly larger than zero, as in our application.

Table 6. The parameter estimates of a GLARMA(2,0) model fitted to the COPD cases.

Variable	Estimates	Standard Error	*p*-Value
β_1	−0.0399	0.0105	0.0001
β_2	0.0347	0.0151	0.0221
β_3	−0.0560	0.0185	0.0025
β_4	0.0538	0.0173	0.0019
β_5	−0.0832	0.0206	0.0000
β_6	0.9277	0.0424	0.0000
β_7	−0.6710	0.0367	0.0000
β_8	−0.3772	0.0188	0.0000
β_9	−0.0695	0.0215	0.0012
β_{10}	0.0623	0.0005	0.0000
ϕ_1	0.0021	0.0017	0.0587
ϕ_2	0.1021	0.0017	0.0000

Algorithm 5 Bootstrap Confidence Interval Calculations for Pollutant Effects

Input: Response variable (monthly COPD cases),
Covariates (pollutants and components of trend and seasonality).
Output: Bootstrap confidence intervals for the pollutant coefficients $\hat{\beta}_j$.
Step 1: Fit a VAR model to the pollutant time series data.
Step 2: Apply PCA to the residuals obtained from the VAR model to extract the principal components.
Step 3: Fit a GLARMA Poisson model with the monthly COPD cases as the response variable and the covariates being the components from PCA, trend, and seasonality components of the COPD cases. Obtain the estimated parameters for each pollutant from the factor loadings (eigenvectors) associated with the eigenvalues of the PCA:

$$\hat{\beta}_j = \sum_{i=1}^{r} a_{ji} v_i, \quad j = 1, \ldots, q$$

where v_i are the estimated coefficients of the i-th principal component and a_{ji} are the eigenvectors.
Step 4:
for $b = 1$ **to** B **do**
Generate bootstrap samples of the response variable Y^* using Algorithm 3
Fit a GLARMA Poisson model with the generated Y^* and the covariates being the components from PCA, trend, and seasonality components of the COPD cases.
Compute the bootstrap sample of the coefficient $\hat{\beta}_j^*$ using the formula:

$$\hat{\beta}_j^* = \sum_{i=1}^{r} a_{ji}^* v_i^*, \quad j = 1, \ldots, q$$

where v_i^* are the resampled estimated coefficients of the i-th principal component and a_{ji}^* are the resampled eigenvectors.
end for
Step 5: Construct confidence intervals for the estimated coefficients, based on the bootstrap samples generated in Step 5, following the procedure described in Algorithm 4.
Step 6: Calculate the average limits of the confidence intervals by obtaining the mean of the lower and upper bounds of each pollutant's parameters.
return Bootstrap confidence intervals for the pollutant coefficients $\hat{\beta}_j$.

Table 7 shows that the 95% CIs calculated using asymptotic and bootstrap techniques were similar, corresponding to the conclusions obtained in the simulation study. This real data analysis is equivalent to scenario 2, once the PCA covariates originated from the VAR process are not autocorrelated. For the parameter ϕ, the numerical study in Section 3.1.1

revealed that the confidence intervals provided by INAR bootstrap and the asymptotic approach are quite close (see Table 2(b)).

Table 7. A comparison of the relative risk and 95% confidence intervals for an interquartile variation of the pollutant concentrations.

	$\widehat{RR}$	Asymptotic CI	INAR(1) Bootstrap CI
CO	0.9677	[0.9425; 0.9936]	[0.9372; 0.9933]
PM_{10}	1.0473	[1.0132; 1.0825]	[1.0174; 1.0783]
NO	0.9466	[0.9121; 0.9825]	[0.9172; 0.9772]
NO_2	1.1294	[1.0804; 1.1806]	[1.0721; 1.1876]
O_3	1.0442	[1.0078; 1.0819]	[1.0068; 1.0802]

4. Discussion

This work proposed to study three bootstrap confidence interval approaches for the relative risk calculated from GLARMA models: the classic model-based approach, a procedure based on the specifications of the model, the sieve bootstrap well known in the literature for real-valued processes, and the recently proposed INAR(1) bootstrap based on the structure of the integer-valued autoregressive process.

An extensive numerical study was performed considering different scenarios and sample sizes. For large samples, the analysis showed that the model parameter (ϕ) could impact the estimates and, consequently, the coverage rate of the intervals. The primary reason for this behavior may be that, for high values of ϕ, the convergence of the maximization method is not guaranteed. Similar behavior has been observed in GLARMA models by [49,50]. In particular, ref. [49] reported that when $\phi = 0.7$, the likelihood function struggles to reach the global maximum. This effect became more pronounced as the complexity of the covariate time structure increased.

The observations in this study agree with the conclusions of [7,8], which demonstrated that removing the temporal structure of covariates prior to modeling improves the accuracy of parameter estimation and confidence interval coverage. In this study, the best results were verified when X_1 did not exhibit time correlation, equivalent to the variables filtered by the VAR process. This reinforces the importance of preprocessing time-dependent covariates in regression models.

Moreover, this study aimed to evaluate the performance of different methodologies for estimating confidence intervals in small-sample scenarios, particularly those mimicking real-world complexities in air pollution data. In all cases, the INAR(1) bootstrap outperforms both the classic model-based and sieve bootstrap approaches. Among the methods evaluated, the INAR(1) bootstrap consistently provided confidence intervals with coverage rates closest to the nominal 95% level, whereas the asymptotic approach tended to underestimate variability. This finding aligns with the observations of [51], which highlighted that using residuals from Poisson regression models typically underestimates the true serial dependence. Thus, for practical applications involving count time series, the INAR(1) bootstrap presents itself as the most reliable alternative, particularly in small-sample contexts.

In a real data set analysis, we assessed the impact of air pollutants on the monthly number of COPD cases in Belo Horizonte, Brazil. The best-fitting model incorporated an autoregressive structure of order 2, effectively capturing the temporal dependence in the data. All estimated coefficients were statistically significant at the 5% significance level, reinforcing the model's reliability and highlighting the importance of the selected predictors—specifically, air pollutants, trend, and seasonality components—in explaining the outcome variable.

To evaluate the robustness of our methodology, we compared the confidence intervals obtained via the INAR(1) bootstrap and the asymptotic approach. As observed in the empirical study, both approaches yielded comparable intervals, confirming the consistency of our method in real-world applications. These findings emphasize that, in practice, the INAR(1) bootstrap provides a reliable alternative for modeling relative risk in environmental epidemiology studies, ensuring more accurate inference even in complex data structures.

Author Contributions: Methodology, A.J.A.C., V.A.R. and G.C.F.; Validation, A.J.A.C.; Formal analysis, A.J.A.C.; Data curation, A.J.A.C.; Writing—original draft, A.J.A.C.; Writing—review & editing, V.A.R., G.C.F. and P.B.; Supervision, V.A.R. and P.B.; Funding acquisition, P.B. All authors have read and agreed to the published version of the manuscript.

Funding: 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 the DATAIA convergence institute as part of the "Programme d'Investissement d'Avenir", (ANR-17-CONV-0003) operated by CentraleSupélec.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Davis, R.; Fokianos, K.; Holan, S.; Joe, H. Count time series: A methodological review. *J. Am. Stat. Assoc.* **2021**, *116*, 1533–1547. [CrossRef]
2. Ostro, B.; Eskeland, G.; Sánchez, J.; Feyzioglu, T. Air pollution and health effects: A study of medical visits among children in Santiago, Chile. *Environ. Health Perspect.* **1999**, *107*, 69–73. [CrossRef] [PubMed]
3. Schwartz, J. Harvesting and long-term exposure effects in the relationship between air pollution and mortality. *Am. J. Epidemiolology* **2000**, *151*, 440–448. [CrossRef]
4. Chen, R.; Chu, C.; Tan, J.; Cao, J.; Song, W.; Xu, X.; Jiang, C.; Ma, W.; Yang, C.; Chen, B.; et al. Ambient air pollution and hospital admission in Shanghai, China. *J. Hazard. Mater.* **2010**, *181*, 234–240. [CrossRef]
5. Borhan, S.; Motevalian, P.; Ultman, J.; Bascom, R.; Borhan, A. A patient-specific model of reactive air pollutant uptake in proximal airways of the lung: Effect of tracheal deviation. *Appl. Math. Model.* **2021**, *91*, 52–73. [CrossRef]
6. Barbera, E.; Currò, C.; Valenti, G. A hyperbolic model for the effects of urbanization on air pollution. *Appl. Math. Model.* **2010**, *34*, 2192–2202. [CrossRef]
7. Souza, J.; Reisen, V.; Franco, G.; Ispány, M.; Bondon, P.; Santos, J. 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* **2018**, *67*, 453–480. [CrossRef]
8. Ispany, M.; Reisen, V.; Franco, G.; Bondon, P.; Cotta, H.; Prezotti, P.; Serpa, F. On Generalized Additive Models with Dependent Time Series Covariates. In *Time Series Analysis and Forecasting. ITISE 2017. Contributions to Statistics*; Rojas, I., Pomares, H., Valenzuela, O., Eds.; Springer: Cham, Switzerland, 2018.
9. Hastie, T.; Tibshirani, R. *Generalized Additive Models*; Chapman and Hall: London, UK, 1990.
10. Liu, B.; Yu, X.; Chen, J.; Wang, Q. Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model. *Atmos. Pollut. Res.* **2021**, *12*, 101144. [CrossRef]
11. Ghaderpour, E.; Pagiatakis, S.D.; Mugnozza, G.S.; Mazzanti, P. On the stochastic significance of peaks in the least-squares wavelet spectrogram and an application in GNSS time series analysis. *Signal Process.* **2024**, *223*, 109581. [CrossRef]
12. Davis, R.; Dunsmuir, W.; Streett, S. Observation driven models for Poisson counts. *Biometrika* **2003**, *90*, 777–790. [CrossRef]
13. Nelder, J.; Wedderburn, R. Generalized linear model. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 370–384. [CrossRef]
14. Rydberg, T.; Shephard, N. Dynamics of trade-by-trade price movements:decomposition and models. *J. Financ. Econ.* **2003**, *1*, 2–25.

15. Jung, R.C.; Kukuk, M.; Liesenfeld, R. Time series of count data: Modelling, estimation and diagnostics. *Comput. Stat. Data Anal.* **2006**, *51*, 2350–2364. [CrossRef]
16. Jung, R.C.; Tremayne, A.R. Useful models for time series of counts or simply wrong ones? *Adv. Stat. Anal.* **2011**, *95*, 59–91. [CrossRef]
17. Karami, S.; Karami, M.; Roshanaei, G.; Farsan, H. Association Between Increased Air Pollution and Mortality from Respiratory and Cardiac Diseases in Tehran: Application of the Glarma Model. *Iran. J. Epidemiol.* **2017**, *12*, 36–43.
18. Ballesteros-Cánovas, J.; Trappmann, D.; Madrigal-González, J.; Eckert, N.; Stoffel, M. Climate warming enhances snow avalanche risk in the Western Himalayas. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3410–3415. [CrossRef] [PubMed]
19. Peitzsch, E.; Pederson, G.; Birkeland, K.; Hendrikx, J.; Fagre, D. Climate drivers of large magnitude snow avalanche years in the U.S. northern Rocky Mountains. *Sci. Rep.* **2021**, *11*, 10032. [CrossRef]
20. Zeger, S.L.; Qaqish, B. Markov regression models for time series: A quasi-likelihood approach. *Biometrics* **1988**, *44*, 1019–1031. [CrossRef] [PubMed]
21. Benjamin, M.; Rigby, R.; Stasinopoulos, D. Generalized autoregressive moving average models. *J. Am. Stat. Assoc.* **2003**, *98*, 214–223. [CrossRef]
22. Heinen, A. *Modelling Time Series Count Data: An Autoregressive Conditional Poisson Model*; Munich Personal RePEc Archive; University Library of Munich: Munich Germany, 2003.
23. Fokianos, K.; Tjosthein, J. Log-linear Poisson autoregression. *J. Multivar. Anal.* **2011**, *102*, 563–578. [CrossRef]
24. Camara, A.J.; Franco, G.; Reisen, V.; Bondon, P. Generalized additive model for count time series: An application to quantify the impact of air pollutants on human health. *Pesqui. Oper.* **2021**, *41*, e241120. [CrossRef]
25. Harvey, A.; Fernandes, C. Time series models for count or qualitative observations. *J. Bus. Econ. Stat.* **1989**, *7*, 407–417. [CrossRef]
26. Gamerman, D.; Santos, T.; Franco, G. A non-Gaussian family of state-space models with exact marginal likelihood. *J. Time Ser. Anal.* **2013**, *34*, 625–645. [CrossRef]
27. Efron, B. Bootstrap methods: Another look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [CrossRef]
28. Freedman, D. On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Stat.* **1984**, *12*, 827–842. [CrossRef]
29. Efron, B.; Tibishirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1986**, *1*, 54–75. [CrossRef]
30. Franke, J.; Kreiss, J. Bootstrapping stationary autoregressive moving average models. *J. Time Ser. Anal.* **1992**, *13*, 297–317.
31. Härdle, W.; Huet, S.; Mammen, E.; Sperlich, S. Bootstrap Inference in Semiparametric Generalized Additive Models. *Econ. Theory* **2004**, *20*, 265–300. [CrossRef]
32. Künsch, H. The Jackknife and the bootstrap for genreal stationary observations. *Ann. Stat.* **1989**, *17*, 1217–1241. [CrossRef]
33. Bühlmann, P. Sieve bootstrap for time series. *Bernoulli* **1997**, *3*, 123–148. [CrossRef]
34. Jentsch, C.; Weiss, C. Bootstrapping INAR models. *Bernoulli* **2019**, *25*, 2359–2408. [CrossRef]
35. Cardinal, M.; Roy, R.; Lambert, J. On the application of integer-valued time series models for the analysis of disease incidence. *Stat. Med.* **1999**, *18*, 2025–2039. [CrossRef]
36. Kim, H.Y.; Park, Y. Bootstrap confidence intervals for the INAR(p) process. *Korean Commun. Stat.* **2006**, *13*, 343–358. [CrossRef]
37. Kim, H.Y.; Park, Y. A non-stationary integer-valued autoregressive model. *Stat. Pap.* **2008**, *49*, 485–502. [CrossRef]
38. Dunsmuir, W. *Handbook of Discrete-Valued Time Series, Handbook of Modern Statistical Methods*; CRC Press: London, UK, 2016.
39. Dunsmuir, W.; Scott, D. The glarma Package for Observation-Driven time series regression of counts. *J. Stat. Softw.* **2015**, *67*, 1–36. [CrossRef]
40. Baxter, L.; Finch, S.; Lipfert, F.; Yu, Q. Comparing estimates of the effects of air pollution on human mortality obtained using different regression methodologies. *Risk Anal.* **1997**, *17*, 273–278. [CrossRef]
41. Kreiss, J.P.; Paparoditis, E.; Politis, D. On the range of validity of the autoregressive sieve bootstrap. *Ann. Stat.* **2011**, *39*, 2103–2130. [CrossRef]
42. McKenzie, E. Some simple models for discrete variate time series. *J. Am. Water Resour. Assoc.* **1985**, *21*, 645–650. [CrossRef]
43. Al-Osh, M.; Alzaid, A. First-order integer-valued autoregressive (INAR(1)) processs. *J. Time Ser. Anal.* **1988**, *8*, 261–275. [CrossRef]
44. Alzaid, A.; Al-Osh, M. An integer-valued pth order autoregressive structure (INAR(p)) process. *J. Appl. Probab.* **1990**, *27*, 314–324. [CrossRef]
45. Du, J.G.; Li, Y. The integer valued autoregressive (INAR(p)) model. *J. Time Ser. Anal.* **1991**, *12*, 129–142.
46. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman and Hall: New York, NY, USA, 1993.
47. Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **1901**, *2*, 559–572. [CrossRef]
48. Zamprogno, B.; Reisen, V.; Bondon, P.; Cotta, H.; Reis, N., Jr. Principal component analysis with autocorrelated data. *J. Stat. Comput. Simul.* **2020**, *90*, 2117–2135. [CrossRef]
49. Franco, G.; Migon, H.; Prates, M. Time series of count data: A review, empirical comparisons and data analysis. *Braz. J. Probab. Stat.* **2019**, *33*, 756–781. [CrossRef]

50. Maia, G.; Franco, G. Conditional parametric bootstrap in GLARMA models. *J. Stat. Comput. Simul.* **2024**, *95*, 330–350. [CrossRef]
51. Davis, R.; Wang, Y.; Dunsmuir, W. *Modelling Time Series of Count Data*; Asymptotics, Nonparametrics, and Time Series; Ghosh, S., Ed.; CRC Press: New York, NY, USA, 1999.