Statistical Papers (2024) 65:5145–5166
https://doi.org/10.1007/s00362-024-01590-w

REGULAR ARTICLE

Additive partial linear models with autoregressive symmetric errors and its application to the hospitalizations for respiratory diseases

Shu Wei Chou-Chen[1] · Rodrigo A. Oliveira[2] · Irina Raicher[3,4] · Gilberto A. Paula[5]

Received: 13 November 2023 / Revised: 5 June 2024 / Published online: 9 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Additive partial linear models with symmetric autoregressive errors of order p are proposed in this paper for modeling time series data. Specifically, we apply this model class to explain the weekly hospitalization for respiratory diseases in Sorocaba, São Paulo, Brazil, by incorporating climate and pollution as covariates, trend and seasonality. The main feature of this model class is its capability of considering a set of explanatory variables with linear and nonlinear structures, which allows, for example, to model jointly trend and seasonality of a time series with additive functions for the nonlinear explanatory variables and a predictor to accommodate discrete and linear explanatory variables. Additionally, the conditional symmetric errors allow the possibility of fitting data with high correlation order, as well as error distributions with heavier or lighter tails than the normal ones. We present the model class and a novel iterative process is derived by combining a P-GAM type algorithm with a quasi-Newton procedure for the parameter estimation. The inferential results, diagnostic procedures, including conditional quantile residual analysis and local influence anal-

✉ Shu Wei Chou-Chen
shuwei.chou@ucr.ac.cr

Rodrigo A. Oliveira
rodrigo.sef@hotmail.com

Irina Raicher
irina.raicher@einstein.br

Gilberto A. Paula
giapaula@ime.usp.br

1 Escuela de Estadística - Centro de Investigación en Matemática Pura y Aplicada, Universidad de Costa Rica, San José, Costa Rica

2 Tribunal Regional do Trabalho da 18ª Região, Goiânia, Goiás, Brazil

3 Department of Neurology, Clinics Hospital of the University of São Paulo Medical School, São Paulo, Brazil

4 Pathology Laboratory, Hospital Israelita Albert Einstein, São Paulo, Brazil

5 Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil

ysis for sensitivity, are discussed. Simulation studies are performed to assess finite sample properties of parametric and nonparametric estimators. Finally, the data set analysis and concluding remarks are given.

Keywords Cubic splines · Cyclic splines · Robust estimation · Penalized likelihood · Climate · Hospitalization

Mathematics Subject Classification 62G08 · 62J05 · 62J20

1 Introduction

In this paper we propose additive partial linear models with p-order autoregressive symmetric errors for modeling time series, which may include linear and nonlinear structures of explanatory variable sets as well as the conditional symmetric errors up to order p. This model class generalizes the previous ones proposed by Relvas and Paula (2016) and Oliveira and Paula (2021).

An advantage of this model class is the possibility of including linear and additive terms in the systematic component of the model. For example to assess the effect of covariates on the mean response by controlling the trend and seasonality of the time series. In addition, the assumption of symmetric error distributions allows kurtosis flexibility and consequently the application of a wide class of error distributions with heavier and lighter tails than the ones of the normal distribution. The P-GAM iterative process is being proposed in this paper for estimating the parametric and additive components in a simpler way than the traditional backfitting algorithm. However, when the number of nonlinear explanatory variables is large, a more appropriate approach rather than an additive function for each nonlinear variable is to accommodate all explanatory variables into the same predictor, named single-index term, that is modeled by a unique additive function. Since the paper by Yu and Ruppert (2002) a wide variety of papers has been published on the single-index approach. In the context of time series one has, for instance, the proposal of a novel partial-linear single-index model with autocorrelated errors (Huang et al. 2016, 2019) and studies on partially linear single-index spatial autoregressive models (Cheng et al. 2019).

Time series data on air pollution, weather conditions and measures of health outcomes (e.g., mortality, hospital admissions) have been modeled to assess how external (pollution, environmental) factors may contribute to increase in morbidity (Bhaskaran et al. 2013). To date, numerous time series analysis have been showing that extreme weather such as, low or high temperatures, relative humidity and air pollution measures are associated with increased risks for a variety of health outcomes (Basu and Samet 2002; Basu 2009; Ye et al. 2012). Respiratory diseases have been representing a huge burden on primary health resources (Lim et al. 2023). COVID-19 studies have showed associations between temperature, relative humidity and stability and transmissibility. Flexible time series model may contribute with the appropriate public health mitigation for outbreak readiness and guide healthcare resources planning.

The inclusion of linear covariates in the proposed model could be beneficial for understanding the linear effects of external factors, in addition to temporal information such as trend and seasonality.

We illustrate in this paper the application of the proposed model class for analyzing the weekly data of respiratory disease hospitalizations associated with climate and air pollution in the city of Sorocaba, São Paulo, Brazil.

The paper is organized as follows. Section 2 presents the model class, some of its properties and the derivation of the penalized likelihood function. In Sect. 3 a P-GAM type iterative process for the estimation of the parametric and nonparametric components is derived that is combined with a quasi-Newton procedure for the correlation structure estimation. Discussion on the derivation of the approximate standard errors, simultaneous confidence intervals and effective degrees of freedom are given. Section 4 describes diagnostic procedures, including residual analysis and sensitivity studies based on local influence. Simulation studies for evaluating the small and large sample behavior of the proposed methodology are described in Sect. 5. The proposed methodology is applied in Sect. 6 for analyzing the data set on weekly hospitalization. Finally, concluding remarks are given in the last section whereas various technical results are given as supplementary materials.

2 The model

Let $y_1, ..., y_n$ be the arranged responses in time. We propose the following additive partial linear model to explain the mean response:

$$\begin{aligned} y_i &= \mathbf{x}_i^\top \boldsymbol{\beta} + f_1(t_{i1}) + \ldots + f_k(t_{ik}) + \epsilon_i, \\ \epsilon_i &= \rho_1 \epsilon_{i-1} + \rho_2 \epsilon_{i-2} + \ldots + \rho_p \epsilon_{i-p} + e_i, \end{aligned} \tag{1}$$

where $\mathbf{x}_i = (1, x_{i2}, \ldots, x_{ir_0})^\top$ contains the intercept and $r_0 - 1$ explanatory variable values, $\boldsymbol{\beta}$ denotes a p-dimensional vector with the coefficients of the parametric component, $f_1(t_{i1}), \ldots, f_k(t_{ik})$ are smoothing functions, that may represent either trend or seasonality, $t_{i1}, \ldots, t_{ik}$ denote time units, for example, daily, weekly or monthly data, whereas the error term is correlated up to p lag, that is $\rho_1, \ldots, \rho_p$ are the autoregressive coefficients and e_i are iid zero mean symmetric errors of dispersion parameter ϕ, that is, $e_i \sim S(0, \phi)$, for $i = 1, \ldots, n$. Model class (1) will be named additive partial linear model with autoregressive symmetric errors of order p.

The probability density function of a symmetric random variable e_i is given by

$$h_e(e_i) = \frac{1}{\sqrt{\phi}} g(\delta_i), \quad e_i \in \mathcal{R},$$

with $\delta_i = \phi^{-1} e_i^2$, and $g : \mathcal{R} \to [0, \infty)$, known as a density generator, satisfying $\int_0^\infty u^{-\frac{1}{2}} g(u) du = 1$. When they exist, we have $\mathrm{E}(e_i) = 0$ and $\mathrm{Var}(e_i) = \xi\phi$, with $\xi > 0$ being a constant that is obtained from the expected value or from the characteristic function. This class includes some well known symmetric distributions, including

normal, Student-t, logistic I and II and power exponential (Fang and Anderson 1990; Cysneiros and Paula 2005, see, for instance).

We propose the use of cubic or cyclic splines (Wood 2017) to approximate the smooth functions $f_1(t_{i1}), ..., f_k(t_{ik})$ with fixed knots $t^0_{1\ell} < \cdots < t^0_{m_\ell \ell}$, $\ell = 1, \ldots, k$, respectively. In this way, model (1) may be written in matrix notation as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}_1\boldsymbol{\gamma}_1 + \ldots + \mathbf{N}_k\boldsymbol{\gamma}_k + \boldsymbol{\epsilon}, \tag{2}$$

where $\mathbf{y}$ is n-dimensional vector of response variables, $\mathbf{X}$ denotes the $(n \times r_0)$ model matrix with rows $\mathbf{x}_i^\top$ (intercept and $r_0 - 1$ covariates) and parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{r_0})^\top$, $\mathbf{N}_\ell$ are $(n \times r_\ell)$ base function matrices with rows $\boldsymbol{\eta}_{i\ell}(t_{i\ell}) = \big(\eta_{1\ell}(t_{i\ell}), \ldots, \eta_{r_\ell \ell}(t_{i\ell})\big)^\top$, whose elements depend on the respective knots, whereas $\boldsymbol{\gamma}_\ell = (\gamma_{1\ell}, \ldots, \gamma_{r_\ell \ell})^\top$ are the parameter vectors, for $i = 1, \ldots, n$ and $l = 1, \ldots, k$.

Therefore, one has the model matrices

$$\mathbf{X} = \begin{bmatrix} 1 & x_{12} & \ldots & x_{1r_0} \\ 1 & x_{22} & \ldots & x_{2r_0} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \ldots & x_{nr_0} \end{bmatrix} \quad \text{e} \quad \mathbf{N}_\ell = \begin{bmatrix} \eta_{1\ell}(t_{1\ell}) & \ldots & \eta_{r_\ell \ell}(t_{1r_\ell}) \\ \eta_{1\ell}(t_{2\ell}) & \ldots & \eta_{r_\ell \ell}(t_{2r_\ell}) \\ \vdots & \ddots & \vdots \\ \eta_{1\ell}(t_{n\ell}) & \ldots & \eta_{r_\ell \ell}(t_{nr_\ell}) \end{bmatrix},$$

where $l = 1, 2, \ldots, k$.

Some properties for the model (1) can be derived.

1. For the first observation, $y_1 \sim S(\mu_1^*, \phi)$ one has that

$$\begin{aligned} \mu_1^* = \mathrm{E}(y_1) &= f_1(t_{11}) + \cdots + f_k(t_{1k}) \\ &= \mathbf{x}_1^\top \boldsymbol{\beta} + \boldsymbol{\eta}_{11}^\top(t_{11})\boldsymbol{\gamma}_1 + \cdots + \boldsymbol{\eta}_{1k}^\top(t_{1k})\boldsymbol{\gamma}_k. \end{aligned}$$

2. For the conditional distribution $y_i \mid y_{i-1}, \ldots, y_{\max(i-p,1)} \sim S(\mu_i^*, \phi)$ it follows

$$\begin{aligned} \mu_i^* &= \mathrm{E}(y_i \mid y_{i-1}, \ldots, y_{\max(i-p,1)}) \\ &= \mathbf{x}_i^\top \boldsymbol{\beta} + \boldsymbol{\eta}_{i1}^\top(t_{i1})\boldsymbol{\gamma}_1 + \cdots + \boldsymbol{\eta}_{ik}^\top(t_{ik})\boldsymbol{\gamma}_k + \sum_{j=1}^{\min(i-1,p)} \rho_j \{y_{i-j} - \mathbf{x}_{(i-j)}^\top \boldsymbol{\beta} \\ &\quad - \boldsymbol{\eta}_{(i-j)1}^\top(t_{(i-j)1})\boldsymbol{\gamma}_1 - \cdots - \boldsymbol{\eta}_{(i-j)k}^\top(t_{(i-j)k})\boldsymbol{\gamma}_j\}, \end{aligned}$$

   for $i = 2, \ldots, n$.
3. In general, the marginal mean is given by

$$\mu_i = \mathrm{E}(y_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \boldsymbol{\eta}_{i1}^\top(t_{i1})\boldsymbol{\gamma}_1 + \cdots + \boldsymbol{\eta}_{ik}^\top(t_{ik})\boldsymbol{\gamma}_k, \quad i = 1, \ldots, n.$$

4. The marginal variance of the model with AR(p) error may be expressed as

$$\mathrm{Var}(y_i) = \xi\phi + \xi\phi \sum_{j=1}^{i-1} \rho_j^2 \mathrm{Var}(y_{i-j}) + 2 \sum_{\substack{j=1 \\ j \neq j' j < j'}}^{i-1} \sum_{j'=1}^{i-1} \rho_j \rho_{j'} \mathrm{Cov}(y_{i-j}, y_{i-j'}).$$

5. The covariances between two observations assume the forms

   (a) For $j = i - 1$,

$$\mathrm{Cov}(y_i, y_{i-j}) = \mathrm{Cov}(y_i, y_1) = \sum_{l=1}^{p} \rho_l \mathrm{E}(y_{i-1} e_1).$$

   (b) For $j = i - 2$,

$$\mathrm{Cov}(y_i, y_{i-j}) = \mathrm{Cov}(y_i, y_2) = \rho_1^2 \mathrm{Cov}(y_{i-1}, y_{i-j-l}) + \sum_{l=1}^{p} \rho_l \mathrm{E}(y_{i-l} e_k) + \rho_1 \sum_{l=2}^{p} \rho_l \mathrm{Cov}(y_{i-l}, y_1).$$

   (c) For $j = i - p$,

$$\mathrm{Cov}(y_i, y_{i-j}) = \mathrm{Cov}(y_i, y_p) = \sum_{l=1}^{p-1} \rho_l^2 \mathrm{Cov}(y_{i-l}, y_{k-l}) + \sum_{l=1}^{p} \rho_l \mathrm{E}(y_{i-l} e_k) + \sum_{\substack{l=1 \\ l \neq l'}}^{p} \sum_{l'=1}^{p-1} \rho_l \rho_{l'} \mathrm{Cov}(y_{i-l}, y_{k-l'}).$$

   (d) For $j < i - p = k$,

$$\mathrm{Cov}(y_i, y_{i-j}) = \mathrm{Cov}(y_i, y_k) = \sum_{l=1}^{p} \rho_l^2 \mathrm{Cov}(y_{i-l}, y_{k-l}) + \sum_{l=1}^{p} \rho_l \mathrm{E}(y_{i-l} e_k) + \sum_{\substack{l=1 \\ l \neq l'}}^{p} \sum_{l'=1}^{p} \rho_l \rho_{l'} \mathrm{Cov}(y_{i-l}, y_{k-l'}).$$

The proposed model class may be applied to study at least three practical situations. First, when there are nonlinear relationships between the response and continuous explanatory variables, that may be modeled by additive functions. Second, when the interest is on the study of the explanatory variable effects in a parametric way, but controlling trend and seasonality of the time series by additive functions. Third, when

the interest is to assess the trend of the time series by some additive function controlling the seasonality by cyclic additive functions. So, since the proposed model class assumes p-order autoregressive errors with kurtosis flexibility, one has a wide class of symmetric models to analyze data sets from these practical situations.

2.1 Penalized likelihood function

By defining $\boldsymbol{\theta} = (\boldsymbol{\gamma}_0^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \phi, \rho_1, \ldots, \rho_p)^\top \in \boldsymbol{\Theta} \subseteq \mathcal{R}^r$, where $\boldsymbol{\gamma}_0 = \boldsymbol{\beta}$ with $r = r_0 + r_1 + \ldots + r_k + p + 1$ the number of estimated parameters, the log-likelihood function of $\boldsymbol{\theta}$ is given by

$$\mathrm{L}(\boldsymbol{\theta}) = -\frac{n}{2}\log(\phi) + \sum_{i=1}^{n} \log\{g(\delta_i)\}. \tag{3}$$

Since maximizing (3) without restrictions on the nonparametric functions can lead to overfitting and unidentifiable parameters, a penalty function for each nonparametric component must be incorporated. By assuming these functions continuous and their second derivative integrable, the penalized log-likelihood function is defined by

$$\mathrm{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathrm{L}(\boldsymbol{\theta}) - \frac{\lambda_1}{2}\int_{a_1}^{b_1} \left[f_1''(t)\right]^2 dt - \ldots - \frac{\lambda_k}{2}\int_{a_k}^{b_k} \left[f_k''(t)\right]^2 dt, \tag{4}$$

which includes penalties of second derivative for $a_\ell \leq t \leq b_\ell$, with a_ℓ and b_ℓ defined according to the domain of each ℓ's smoothing function, and $\lambda_\ell > 0$ being the smoothing parameters that are estimated separately, $\ell = 1, \ldots, k$.

It may be showed that

$$\int_{a_\ell}^{b_\ell} \left[f_\ell''(t)\right]^2 dt = \boldsymbol{\gamma}_\ell^\top \mathbf{M}_\ell \boldsymbol{\gamma}_\ell,$$

where $\mathbf{M}_\ell = \mathbf{D}_\ell^\top \mathbf{B}_\ell^{-1} \mathbf{D}_\ell$ are non-negative definite matrices of dimensions $(r_l \times r_\ell)$, named penalty matrices (Lancaster and Salkauskas 1986; Wood 2017), $\ell = 1, \ldots, k$, which are defined according to the knots (see Table 5.1 in Wood 2017). Consequently, the penalized log-likelihood function can be expressed by

$$\mathrm{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathrm{L}(\boldsymbol{\theta}) - \frac{\lambda_1}{2}\boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \ldots - \frac{\lambda_k}{2}\boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k, \tag{5}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)^\top$ are the smoothing parametric vector. In the following section, a joint iterative process, combining a P-GAM type iterative process and a quasi-Newton algorithm, is proposed for obtaining the maximum penalized likelihood estimate (MPLE) $\widehat{\boldsymbol{\theta}}$ by maximizing (5) for fixed $\boldsymbol{\lambda}$. The Sect. 3.3 describes two methods to optimize the parameters $\boldsymbol{\lambda}$ in order to select an appropriate smoothing parameter.

3 Parameter estimation

The traditional and elegant iterative process for deriving the MPLE of $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0^\top, \ldots, \boldsymbol{\gamma}_k^\top)^\top$ is the backfitting (Gauss–Seidel) algorithm that has been largely applied in various classes of additive models (Hastie and Tibshirani 1990; Green and Silverman 1994, see, for instance). However, this algorithm may get slow as the number of additive components increases and particularly for correlated data, such as time series. So, similarly to Cardozo et al. (2022) we develop a joint iterative process, that combines the P-GAM algorithm proposed by Marx and Eilers (1998) with a quasi-Newton algorithm, for obtaining the MPLE $\widehat{\boldsymbol{\theta}}$.

From Section S.1 of Supplementary Materials (SM), the score function for $\boldsymbol{\gamma}$ may be expressed as

$$\mathbf{U}_p^{\gamma} = \phi^{-1}\mathbf{N}_A^\top \mathbf{D}_v(\mathbf{A}\mathbf{y} - \mathbf{N}_A\boldsymbol{\gamma}) - \mathbf{M}_{\gamma}(\boldsymbol{\lambda})\boldsymbol{\gamma},$$

where $\mathbf{N}_A = (\mathbf{A}\mathbf{N}_0, \mathbf{A}\mathbf{N}_1, \ldots, \mathbf{A}\mathbf{N}_k)^\top$, $\mathbf{A}$ is a correlation matrix whereas $\mathbf{D}_v = \text{diag}\{v_1, \ldots, v_n\}$ with $v_i > 0$ being weights that depend on the error distribution and $\mathbf{M}_{\gamma}(\boldsymbol{\lambda}) = \text{blockdiag}\left\{\mathbf{0}_{r_0}, \lambda_1\mathbf{M}_1, \ldots, \lambda_k\mathbf{M}_k\right\}$. Then, by fixing $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$, the profiled MPLE of $\boldsymbol{\gamma}$ may be obtained by setting $\mathbf{U}_p^{\gamma} = \mathbf{0}$ whose solution leads to the following iterative process:

$$\boldsymbol{\gamma}^{(u+1)} = \left\{\mathbf{N}_A^\top \mathbf{D}_v^{(u)}\mathbf{N}_A + \phi\mathbf{M}_{\gamma}(\boldsymbol{\lambda})\right\}^{-1}\mathbf{N}_A^\top \mathbf{D}_v^{(u)}\mathbf{y}_d, \tag{6}$$

for $u = 0, 1, \ldots$, where $\mathbf{y}_d = \mathbf{A}\mathbf{y}$ works as a dependent variable. It is assumed that $\mathbf{N}_A$ is a full column rank matrix, which guarantees the identification of the additive functions in (1). The profiled MPLE of $\boldsymbol{\xi} = (\boldsymbol{\rho}^\top, \phi)^\top$ may be obtained as

$$\boldsymbol{\xi}^{(s+1)} = \underset{(\boldsymbol{\rho},\phi)}{\arg\max}\ \mathrm{L}_p(\boldsymbol{\gamma}^{(s+1)}, \boldsymbol{\rho}, \phi, \boldsymbol{\lambda}) \tag{7}$$

for $s = 0, 1, 2 \ldots$, where $\boldsymbol{\gamma}^{(s+1)}$ corresponds to the profiled MPLE of $\boldsymbol{\gamma}$, obtained at the $(s+1)$th convergence of the iterative process (6). Equation (7) may be solved by applying the quasi-Newton Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (Davidon 1991; Mittelhammer et al. 2000) or the "L-BFGS-B" (Byrd et al. 1995) procedure. The MPLEs $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\xi}}$ are obtained by combining the iterative processes (6) and (7).

Then, by fixing $\boldsymbol{\lambda}$ and defining the matrices $\mathbf{N}_1, \ldots, \mathbf{N}_k$ and $\mathbf{M}_1, \ldots, \mathbf{M}_k$ as well as the respective knots, we propose the following algorithm to obtain $\widehat{\boldsymbol{\theta}}$:

1. Set the counter u to zero and the initial values $\boldsymbol{\gamma}^{(0)}$ and $\boldsymbol{\rho}^{(0)}$.
2. Performing the iterative process (6) for obtaining the profiled MPLE of $\boldsymbol{\gamma}$.
3. Given the profiled MPLE of $\boldsymbol{\gamma}$ obtained from the $(s+1)$th convergence of the iterative process (6), obtaining the profiled MPLE of $\boldsymbol{\xi}$ from (7).
4. Alternating the iterative processes (6) and (7) until the joint convergence for obtaining the MPLE $\widehat{\boldsymbol{\theta}}$.

In particular, for normal error, one has $\mathbf{D}_v = \mathbf{I}_n$ and consequently the profiled MPLE of $\boldsymbol{\gamma}$, given $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$, takes the form $\widehat{\boldsymbol{\gamma}} = \left\{\mathbf{N}_A^\top \mathbf{N}_A + \phi \mathbf{M}_\gamma(\boldsymbol{\lambda})\right\}^{-1} \mathbf{N}_A^\top \mathbf{y}_d$, that should be iterated with (7) until the convergence.

3.1 Inferential results

The asymptotic normality of the MPLE $\widehat{\boldsymbol{\gamma}}$ with variance–covariance matrix $\text{Var}(\widehat{\boldsymbol{\theta}}) = (\mathcal{I}_p^{\theta\theta})^{-1}$, where $\mathcal{I}_p^{\theta\theta}$ denotes the penalized Fisher information matrix for $\boldsymbol{\theta}$, has support in the Bayesian approach for linear models, as pointed by Wood (2017). In addition, simulation studies performed by Oliveira and Paula (2021) for additive models with autoregressive symmetric errors give support for the consistency of the MPLEs $\widehat{\boldsymbol{\rho}}$ and $\widehat{\phi}$ as well as for the MPLEs of the additive functions.

One has orthogonality among the parameters $\boldsymbol{\gamma}$, $\boldsymbol{\rho}$ and ϕ (see Sections S.2 and S.3 from SM). So, the asymptotic variance–covariance matrix for $\widehat{\boldsymbol{\gamma}}$ takes the form

$$\text{Var}(\widehat{\boldsymbol{\gamma}}) = (\mathcal{I}_p^{\gamma\gamma})^{-1} = \left\{\frac{4d_g}{\phi}\mathbf{N}_A^\top \mathbf{N}_A + \mathbf{M}_\gamma(\boldsymbol{\lambda})\right\}^{-1},$$

whereas the asymptotic variance of $\widehat{\phi}$ becomes given by $\text{Var}(\widehat{\phi}) = (\mathcal{I}_p^{\phi\phi})^{-1} = 4\phi^2/[n(4f_g - 1)]$, where d_g and f_g are constants that depend on the symmetric distribution. Similarly, one may derive the asymptotic variance–covariance matrix for $\widehat{\boldsymbol{\rho}}$ as $\text{Var}(\widehat{\boldsymbol{\rho}}) = (\mathcal{I}_p^{\rho\rho})^{-1}$ whose elements are given in Section S.3 (SM). Finally, pointwise asymptotic confidence bands for the additive functions at the knots, namely $(f_\ell(t_{1\ell}^0), \ldots, f_\ell(t_{m_\ell \ell}^0))^\top$ for $\ell = 1, \ldots, k$, may be constructed by using the approach described in Vanegas and Paula (2016).

3.2 Effective degrees of freedom

Given that penalizing the smoothing functions $f_1(t_1)$, ..., $f_k(t_k)$ contributes to a shrinkage of the MPLEs $\widehat{\boldsymbol{\gamma}}_1, \ldots, \widehat{\boldsymbol{\gamma}}_k$ with respect to the unpenalized MLEs, obtaining the effective degrees of freedom according to the MPLEs is essential for model selection and inferential procedures, such as hypothesis testing. The idea, similarly to the linear regression, is to assess the cost of estimating the linear predictor $\widehat{\mathbf{N}}_A\widehat{\boldsymbol{\gamma}}$ from the dependent response variable $\widehat{\mathbf{y}}_d$. Then, since one may express $\widehat{\mathbf{N}}_A\widehat{\boldsymbol{\gamma}} = \widehat{\mathbf{D}}_v^{-\frac{1}{2}}\widehat{\mathbf{H}}(\boldsymbol{\lambda})\widehat{\mathbf{D}}_v^{\frac{1}{2}}\widehat{\mathbf{y}}_d$, where

$$\widehat{\mathbf{H}}(\boldsymbol{\lambda}) = \widehat{\mathbf{D}}_v^{\frac{1}{2}}\widehat{\mathbf{N}}_A \left\{\widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A + \phi \mathbf{M}_\gamma(\boldsymbol{\lambda})\right\}^{-1} \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v^{\frac{1}{2}},$$

the effective degrees of freedom may be estimated as the sum of the eigenvalues of the linear smoother $\widehat{\mathbf{D}}_v^{-\frac{1}{2}}\widehat{\mathbf{H}}(\boldsymbol{\lambda})\widehat{\mathbf{D}}_v^{\frac{1}{2}}$ that corresponds to its trace [see, for instance, Chapter 5 in Hastie and Tibshirani (1990); Green and Silverman (1994) and Wood (2017)]. Thus, the effective degrees of freedom according to $\widehat{\boldsymbol{\gamma}}$ are given by $\text{df}_s(\boldsymbol{\lambda}) = \text{tr}\{\widehat{\mathbf{D}}_v^{-\frac{1}{2}}\widehat{\mathbf{H}}(\boldsymbol{\lambda})\widehat{\mathbf{D}}_v^{\frac{1}{2}}\} = \text{tr}\{\widehat{\mathbf{H}}(\boldsymbol{\lambda})\}$.

One may establish the relationship (Eilers and Marx 1996)

$$\mathrm{df}_s(\boldsymbol{\lambda}) = \sum_{i=1}^{r_0+r_1+\ldots+r_k} \frac{1}{1+\alpha_i(\boldsymbol{\lambda})},$$

with $\alpha_i(\boldsymbol{\lambda}) \geq 0$ the eigenvalues of the non-negative definite matrix $\mathbf{Q}^{-\frac{1}{2}}\mathbf{M}_\gamma(\boldsymbol{\lambda})\mathbf{Q}^{-\frac{1}{2}}$ and $\mathbf{Q}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}} = \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A$, for $i = 1, \ldots, r_0 + r_1 + \ldots + r_k$. The effective degrees of freedom $\mathrm{df}_s(\lambda_0), \mathrm{df}_s(\lambda_1), \ldots, \mathrm{df}_s(\lambda_k)$ estimated for the parameter estimates $\widehat{\boldsymbol{\gamma}}_0, \widehat{\boldsymbol{\gamma}}_1, \ldots, \widehat{\boldsymbol{\gamma}}_k$ correspond, respectively, to the sum of the first r_0, followed by the sum of the next r_1 until the sum of the last r_k eigenvalues of the linear smoother $\widehat{\mathbf{H}}(\boldsymbol{\lambda})$. Consequently, the total effective degrees of freedom of the fitted model is obtained by $\mathrm{df}(\boldsymbol{\lambda}) = \mathrm{df}_s(\boldsymbol{\lambda}) + p + 1$.

3.3 Smoothing parameter

The estimation of the smoothing parameter $\boldsymbol{\lambda}$ may be performed by some information criterion, generalized cross-validation or direct maximization of the penalized log-likelihood function. Thus, for a grid of $\boldsymbol{\lambda}$ values, the MPLE of $\boldsymbol{\theta}$ may be obtained by minimizing either the Akaike or the Schwarz criteria, respectively, defined as

$$\mathrm{AIC}(\boldsymbol{\lambda}) = -2\mathrm{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) + 2\mathrm{df}(\boldsymbol{\lambda}) \;\text{ and }\; \mathrm{BIC}(\boldsymbol{\lambda}) = -2\mathrm{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) + \log(n)\mathrm{df}(\boldsymbol{\lambda}).$$

As another option, the generalized cross-validation method (Wood 2017), which is defined by minimizing the function

$$\mathrm{GCV}(\boldsymbol{\lambda}) = \frac{n\sum_{i=1}^n \{y_i - \widehat{\mathrm{E}}(y_i)\}^2}{\{n - \mathrm{df}(\boldsymbol{\lambda})\}^2}$$

may be applied, as well as the faster criterion of maximizing the function

$$\mathrm{L}_p(\boldsymbol{\gamma}^{(s+1)}, \boldsymbol{\rho}^{(s+1)}, \phi^{s+1)}, \boldsymbol{\lambda})$$

after each cycle of the algorithm (6) and (7). For example, by using the procedure `optim` available in the `R` package.

4 Diagnostics methods

Once the model is selected and fitted, diagnostic procedures are recommended to assess the model adequacy and possible presence of atypical observations.

4.1 Residual analysis

Residual analysis is being applied in statistical modeling to assess important deviations from the model assumptions as well as to identify outliers. Quantile residuals (Dunn and Smyth 1996), originally proposed for assessing goodness of fit in statistical models under independent observations, have been largely applied and may be extended to correlated data, such as in multivariate and time series models, as pointed out by Barros and Paula (2019). Such residuals are derived in these models from the conditional cumulative function distributions of the responses. Specifically, for the additive partial linear model (2), we denote the conditional cumulative distribution functions by $F_{y_1}(y_1;\boldsymbol{\theta})$, $F_{y_2|y_1}(y_2;\boldsymbol{\theta}),\ldots,F_{y_n|(y_1,\ldots,y_{n-1})}(y_n;\boldsymbol{\theta})$ and the conditional quantile residuals by $r_{q_1}=\Phi^{-1}\{F_{y_1}(y_1;\widehat{\boldsymbol{\theta}})\}$, $r_{q_2}=\Phi^{-1}\{F_{y_2|y_1}(y_2;\widehat{\boldsymbol{\theta}})\},\ldots,r_{q_n}=\Phi^{-1}\{F_{y_n|(y_1,\ldots,y_{n-1})}(y_n;\widehat{\boldsymbol{\theta}})\}$, where $\Phi(\cdot)$ is the cumulative distribution function of the N(0, 1). For large n, the conditional quantile residuals, $r_{q_1},\ldots,r_{q_n}$, are independent distributed as the standard normal distribution. So, the qqplot between the conditional quantile residuals and the quantiles of the N(0, 1) may be employed to assess deviations from the model error assumptions. Another usual graph, between the conditional quantile residuals and time, may be also applied to assess the error variability control over time.

4.2 Sensitivity analysis

The aim of sensitivity analysis is to assess the effects of perturbations in the model and/or data, on the parameter estimates. The likelihood displacement $\mathrm{LD}(\boldsymbol{\delta})=2\{\mathrm{L}_p(\widehat{\boldsymbol{\theta}},\boldsymbol{\lambda})-\mathrm{L}_p(\widehat{\boldsymbol{\theta}}_\delta,\boldsymbol{\lambda})\}$, where $\widehat{\boldsymbol{\theta}}_\delta$ is the MPLE under the perturbed penalized log-likelihood function $\mathrm{L}_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\delta})$, with $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_n)^\top$ denoting the perturbation vector, is usually considered as the influence measure in additive partial linear models (see, for instance, Oliveira and Paula 2021).

In the context of linear models with autoregressive symmetric errors, Liu (2000, 2004) presents derivations and applications of the local influence procedure. In the sequel, curvatures of local influence are derived in arbitrary $(n\times 1)$ unitary directions $\|\boldsymbol{\ell}\|=1$ and evaluated at the MPLE $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\delta}_0$ (the no perturbation vector).

The conformal normal curvature of local influence, proposed by Poon and Poon (1999), is defined as

$$\mathrm{C}_{\boldsymbol{\ell}}(\boldsymbol{\theta})=\boldsymbol{\ell}^\top\mathbf{C}\boldsymbol{\ell}/\sqrt{\mathrm{tr}(\mathbf{C}^2)},$$

where $0\le\mathrm{C}_{\boldsymbol{\ell}}(\boldsymbol{\theta})\le 1$, $\mathbf{C}=\boldsymbol{\Delta}^\top\{(-\ddot{\mathbf{L}}_p^{\widehat{\theta\theta}})^{-1}\}\boldsymbol{\Delta}$ with $-\mathbf{L}_p^{\theta\theta}$ being the observed information matrix of $\boldsymbol{\theta}$, $\mathrm{tr}(\mathbf{C}^2)=\sum_{i=1}^n\alpha_i^2$ with $\alpha_1,\ldots,\alpha_n$ being the eigenvalues of $\mathbf{C}$ and $\boldsymbol{\Delta}=(\boldsymbol{\Delta}_0^\top,\boldsymbol{\Delta}_1^\top,\ldots,\boldsymbol{\Delta}_k^\top,\boldsymbol{\Delta}_{k+1}^\top,\boldsymbol{\Delta}_{k+2}^\top,\ldots,\boldsymbol{\Delta}_{k+p+1}^\top)^\top$ is a $(r_0+r_1+\ldots+r_k+p+1\times n)$ matrix that contains sub matrices related with the perturbation schemes (derivation in Section S.4 of SM).

Several diagnostic graphs may be derived here. For instance, by denoting the normalized eigenvalues as $\widehat{\alpha}_{\max}=\widehat{\alpha}_1\ge\cdots\ge\widehat{\alpha}_k\ge q/\sqrt{n}>\widehat{\alpha}_{k+1}\ge\cdots\ge\widehat{\alpha}_n\ge 0$, an

aggregate influence measure for every q influential eigenvector is defined as

$$m(q)_i = \sqrt{\sum_{j=1}^{k} \widehat{\alpha}_j e_{ji}^2},$$

where e_{ji} is the ith component of the jth eigenvector $\mathbf{e}_j$ corresponding to the jth normalized eigenvalue $\widehat{\alpha}_j$, for $i = 1, \ldots, n$, $j = 1, \ldots, k$ and $q = 0, 1, 2, \ldots$. The index plot of $m(q)_i$ is suggested by Poon and Poon to expose those observations that are q-influential, that is, influential for all eigenvectors, so that $\mathrm{C}_{e_j}(\boldsymbol{\theta}) \geq q/\sqrt{n}$.

To be precise, if our focus is on assessing the conformal normal curvature along the direction of the ith observation, denoted by the $(n \times 1)$ vector $\mathbf{d}_i$, which consists of zeros with a value of 1 in the ith position, we have

$$\mathrm{C}_{d_i}(\boldsymbol{\theta}) = \mathrm{C}_i = m^2(0)_i = \sum_{j=1}^{n} \widehat{\alpha}_j e_{ji}^2,$$

which matches to the square of the cumulative contribution of the orthonormal eigenvectors. A cutoff criterion, proposed by Lee and Xu (2004), assesses those observations such that $\mathrm{C}_i > \overline{\mathrm{C}} + c\mathrm{SD(C)}$ to be possible influential, where $\bar{\mathrm{C}}$ and SD(C) denote, respectively, the mean and standard deviation of $\{\mathrm{C}_i, i = 1, \ldots, n\}$ and c is a cutoff value defined appropriately.

For a particular partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ with $\boldsymbol{\theta}_1$ defined as the interest, e.g. $\boldsymbol{\theta}_1 = \boldsymbol{\gamma}$, $\theta_1 = \phi$ or $\theta_1 = \boldsymbol{\rho}$, we may determine the conformal normal curvature as

$$\mathrm{C}_{\boldsymbol{\ell}}(\boldsymbol{\theta}_1) = \boldsymbol{\ell}^\top \mathbf{C}_1 \boldsymbol{\ell} / \sqrt{\mathrm{tr}(\mathbf{C}_1^2)},$$

where $0 \leq \mathrm{C}_{\boldsymbol{\ell}}(\boldsymbol{\theta}_1) \leq 1$, $\mathbf{C}_1 = \boldsymbol{\Delta}^\top \{(-\ddot{\mathbf{L}}_p^{\widehat{\theta}\widehat{\theta}})^{-1} - \mathbf{G}_p^{\widehat{\theta}_2\widehat{\theta}_2}\}\boldsymbol{\Delta}$ with $\mathbf{G}_p^{\theta_2\theta_2} = \mathbf{G}_p^{\theta_2\theta_2} = \mathrm{blockdiag}\{\mathbf{0}, (\ddot{\mathbf{L}}_p^{\theta_2\theta_2})^{-1}\}$. Therefore, the index plot of $\mathbf{C}_i(\boldsymbol{\theta}_1)$ can be performed.

In general, some confirmatory analysis is recommended to verify if the pointed out observations from the sensitivity graphs are, in fact, more influential than the remaining observations. In some regression analysis, for example, the impact of dropping the pointed out observations is compared with the impact of dropping a sample of not highlighted observations. However, this procedure is not usual in time series models. So, the useful of the sensitivity graphs in this work is in the sense of model selection. Such models for which the parameter estimates are more resistant to the perturbation schemes are candidates to be chosen.

5 Simulation studies

In this section, we carry out simulation studies to evaluate the small and large sample behavior of the proposed estimation method of the model (1). The observations are

generated using the model with $r_0 - 1 = 2$ covariates and $k = 2$, that is

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + f_1(t_i) + f_2(t_i) + \epsilon_i, \\ \epsilon_i &= \rho_1 \epsilon_{i-1} + \rho_2 \epsilon_{i-2} + \ldots + \rho_p \epsilon_{i-p} + e_i, \end{aligned} \tag{8}$$

where $e_i \sim S(0, \phi)$, $f_1(t_i) = -30\cos(\pi t_i)$ and $f_2(t_i) = 10\sin(2\pi t_i)$ with $t_i = \frac{t}{T}$, for $i = 1, ..., n$. For the parametric part, we assume $\beta_1 = 5$, $\beta_2 = 3$, $\beta_3 = 2$ and $\phi = 5$. The covariates x_{i1} and x_{i2} are generated independently from a uniform distribution in the interval [0, 3].

For the error structure, normal (N), Student-t with $\nu = 3$ (t_3) and power exponential with light tail (PE_1) and heavy tail (PE_2), that is, with shape parameter $\kappa = -0.3$ and 0.5, respectively, are assumed. Moreover, we consider $p = 1$ with $\rho = -0.6, 0.5$, and $p = 2$ with $(\rho_1, \rho_2) = (-0.5, -0.3)$ and $(0.6, 0.2)$. In total, 16 scenarios are simulated for different sample sizes, namely $n = 100, 500, 1000$, based on $R = 5000$ independent replications.

To assess the estimation properties of the parametric part (β_1, β_2, β_3, ϕ, $\boldsymbol{\rho}$), the Monte Carlo (MC) mean, the bias and the mean squared error (MSE) are computed as bias $= \bar{\hat{\theta}} - \theta$ and MSE $= \sum_{r=1}^{R} (\hat{\theta}^{(r)} - \bar{\hat{\theta}})^2 / R$, with $\bar{\hat{\theta}}$ being the MC mean for a specific parameter θ. On the other hand, the consistency of the additive functions (f_1 and f_2) are evaluated by computing the average estimates $\bar{\hat{f}}_1$ and $\bar{\hat{f}}_2$.

The results are summarized in Section S.5 of the SM. Tables S.1 and S.2 show that the bias and MSE of all estimates, except for ϕ, decrease as the sample size increases, indicating consistency. Specifically for ϕ, the bias does not seem to decrease, while the MSE becomes smaller as n increases. On the other hand, we observe that the average estimates of $\bar{\hat{f}}_1$ and $\bar{\hat{f}}_2$ converge to the true functions f_1 and f_2, respectively (Figs. S.1 and S.2 of SM).

6 Application

We illustrate the proposed model with the application of the weekly count of hospitalization due to respiratory diseases in an epidemiological week, in Sorocaba from January 1st 2005 to November 28, 2022. Figure 1 shows the hospitalization time series and its monthly empirical distribution. A positive trend is visualized in the first figure and both figures show a strong seasonal effects, which high hospitalization from April to August may be due to the winter season. We are interested in modeling the effect of climatic variables on the hospitalization due to respiratory diseases.

6.1 Model selection

In order to incorporate the trend and seasonality into the model, we eliminate the week 53 in some of the years, in order to standardize 52 weeks in each year. In total, we have 935 weeks. Then, we model the data using the model (1) by considering the following explanatory variables:

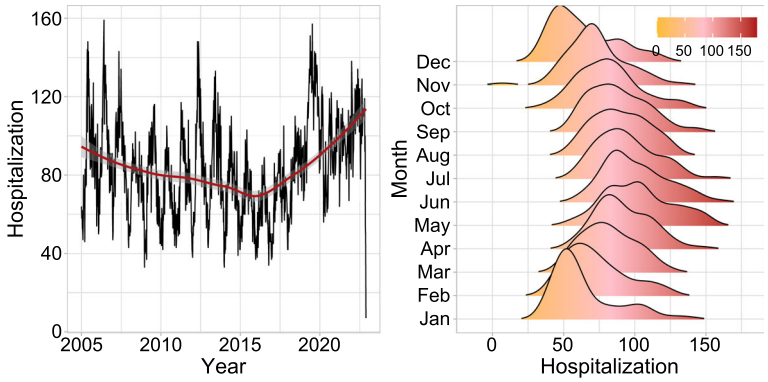


Fig. 1 The time series of respiratory hospitalizations and its empirical distribution in Sorocaba from January 2005 to November 2022

1. $r_0 - 1 = 24$ environment covariates, that include the weekly maximum, minimum and average of the following covariates: particulate matter 10 micrometers or less in diameter (PM10), nitrogen monoxide (NO), nitrogen dioxide (NO_2), nitrogen oxides (NO_x), ozone (O_3), air temperature (TEMP), daily temperature amplitude ($AMPL_{temp}$) and relative humidity (RH).
2. $f_1(\text{time}_i)$ and $f_2(\text{week}_i)$, smooth functions approximated by cubic regression and cyclic cubic regression splines, respectively, in order to model the trend and seasonality of the data.

Following the recommendation from Ruppert et al. (2003), we set $r_1 = \min\{(1/4) \cdot 935, 35\} = 35$ and $r_2 = \min\{(1/4) \cdot 52, 35\} = 13$ equidistant knots for f_1 and f_2, respectively. At this initial stage, we set smoothing parameters as $(\lambda_1, \lambda_2) = (100, 10)$.

Then, we execute the forward stepwise algorithm using 10% of significance level, assuming AR(5) normal error structure, in order to select the most significant covariates in the linear parametric part. The reduced model contains five covariates: weekly average of PM10 ($PM10_{avg}$), minimum of NO_2 ($NO_{2,min}$), maximum of NO_2 ($NO_{2,max}$), average of NO (NO_{avg}) and the minimum of relative humidity (RH_{min}), in order of inclusion. Similar results were obtained under $p \leq 5$ and with Student-t and Power Exponential error models.

In addition, we fix the parametric structure with these five covariates and the additive part and fit the model under three different error structures: Normal, Student-t with $\nu = 5, 10, 15$ and Power Exponential with shape parameter $\kappa = 0.1, 0.3, 0.7$. Moreover, autoregressive error term up to $p = 5$ is considered. Table 1 shows the goodness of fit measures of all combinations. Both AIC and BIC (in bold) indicate that the best model is the Student-t with $\nu = 10$ and AR(4) error structure, described as following:

$$y_i = \beta_1 + \{\text{PM10}_{\text{avg}}\}_i \cdot \beta_2 + \{\text{NO}_{2,\text{min}}\}_i \cdot \beta_3 + \{\text{NO}_{2,\text{max}}\}_i \cdot \beta_4 + \\ + \{\text{NO}_{\text{avg}}\}_i \cdot \beta_5 + \{\text{RH}_{\text{min}}\}_i \cdot \beta_6 + f_1(\text{time}_i) + f_2(\text{week}_i) + \epsilon_i, \quad (9)$$

Table 1 Goodness of fit measures of the model under Normal, Student-t with $\nu = 5, 10, 15$ and Power Exponential with shape parameter $\kappa = 0.1, 0.3, 0.7$ and AR(p) error structure up to $p = 5$, fitted to the hospitalization time series

Information Criteria	AR structure Error	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
AIC	Normal	7528.8660	7451.2774	7434.5763	7430.1617	7432.1606
	t_5	7538.6168	7457.0348	7431.0665	7423.4061	7423.4061
	t_{10}	7523.1181	7444.2018	7421.6561	**7415.3582**	7415.3582
	t_{15}	7521.6953	7443.5118	7422.4434	7416.7664	7416.7664
	$PE_{0.1}$	7525.8424	7446.2160	7427.2506	7422.0210	7422.0210
	$PE_{0.3}$	7530.5659	7446.5169	7423.9540	7417.0719	7417.0719
	$PE_{0.7}$	7563.6926	7471.8848	7445.9465	7435.1780	7435.1780
BIC	Normal	7586.4291	7510.8021	7498.5436	7499.0501	7505.8575
	t_5	7598.9541	7519.0809	7497.3928	7494.7580	7501.4115
	t_{10}	7582.3829	7505.3208	7487.0880	**7485.7545**	7492.5959
	t_{15}	7580.5141	7504.2311	7487.5082	7486.7352	7493.5966
	$PE_{0.1}$	7584.8220	7507.0698	7492.4985	7492.2513	7499.1663
	$PE_{0.3}$	7592.6551	7510.3168	7492.2611	7490.4004	7497.2215
	$PE_{0.7}$	7634.3350	7547.1010	7524.3753	7519.5256	7526.9660

Table 2 Generalized cross validation (GCV) measures for all combinations of $\lambda_1 = (10, 20, 30, 40, 50)$ and $\lambda_2 = (6, 12, 18, 24)$

	λ_2			
λ_1	6	12	18	24
10	**132.8354**	134.4184	135.4307	136.1104
20	133.0057	134.5791	135.5748	136.2445
30	133.0894	134.6639	135.6568	136.3192
40	133.1486	134.7241	135.7148	136.3759
50	133.1948	134.7738	135.7622	136.4171

where $\epsilon_i = \rho_1\epsilon_{i-1} + \rho_2\epsilon_{i-2} + \rho_3\epsilon_{i-3} + \rho_4\epsilon_{i-4} + e_i$ and $e_i \overset{\text{iid}}{\sim} t_{10}$, for $i = 1, \ldots, 935$.

Finally, in order to choose the appropriate smoothing parameters (λ_1, λ_2), we selected the model with the lowest Generalized Cross Validation (GCV) by using a grid of different values for them, $\lambda_1 = (10, 20, 30, 40, 50)$ and $\lambda_2 = (6, 12, 18, 24)$. We selected $(\lambda_1, \lambda_2) = (10, 6)$ (in bold) as the most appropriate smoothing values (Table 2).

6.2 Residual and sensitivity analyses

Before the interpretation of the fitted model, we perform residual and sensitivity analysis in order to evaluate the adequacy of the model and possible presence of atypical observations.

Figure 2 presents various graphs related with the conditional quantile residual, such as r_{q_i} against the time, histogram, density, autocorrelation and partial autocorrelation functions of r_{q_i}, quantile of r_{q_i} against the N(0, 1) quantile and r_{q_i} against the estimated weight $\hat{v}_i$. The graphs indicate no trend and seasonality pattern and that the Student-t with $\nu = 10$ and AR(4) error structure seems to perform a suitable fit. As it is well known in Student-t error models the iterative process assigns smaller weights for larger residuals. The absence of autocorrelation of the residuals up to 60 lags is not rejected by the Ljung-Box test.

In addition, to assess the sensitivity of the parameter estimates under three different perturbation schemes, we perform index plots of C_i for assessing the local influence on the MPLEs. Figure 3 presents the index plots of C_i for assessing the local influence on $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{\phi}$ and $\widehat{\boldsymbol{\rho}}$ under the case-weight perturbation scheme (left column) and under the response perturbation scheme (right column). Similarly, Fig. 4 presents the index plots of C_i for assessing the local influence on the MPLEs under each explanatory variable perturbation scheme. Since one has a large sample, the cutoff value $c = 4$ is considered in each graph. So, only observations suspect to be highly influential on the MPLEs are pointed out.

From all graphs few points are highlighted, which are identified as follows:

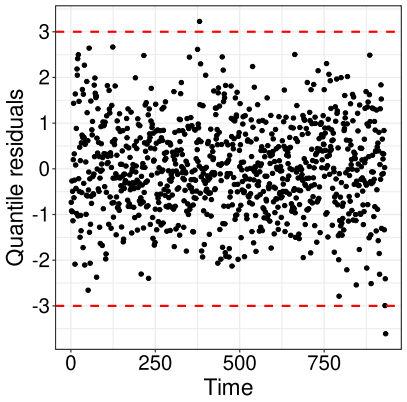
1. Under the case-weight perturbation scheme, possible influential periods are week 1 in 2005, week 20 in 2012 and weeks 5, 47 and 48 in 2022.
2. Week 1 in 2005 and week 26 in 2014 are pointed out as possible influential periods under the response perturbation scheme.
3. Finally, under explanatory variable perturbation schemes, week 1 in 2005, week 41 in 2008, week 24 in 2010, weeks 33 and 40 in 2012 and week 26 in 2014 are highlighted.

A more detailed analysis should be performed to understand the reasons that lead such periods to be more sensitive on the MPLEs. However, as mentioned before, the aim of the sensitivity studies in this work is help in choosing of appropriate error models whose MPLEs are resistant to perturbation schemes.

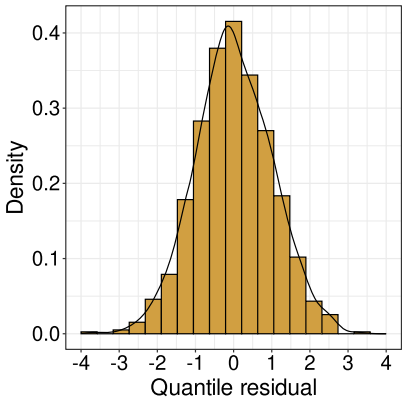
6.3 Model interpretation

In this section we perform a model interpretation, particularly the parameter estimates from the parametric component and the additive trends. Table 3 presents the parameter estimates from the parametric component and one may obtain similar interpretations of linear models, as follows:

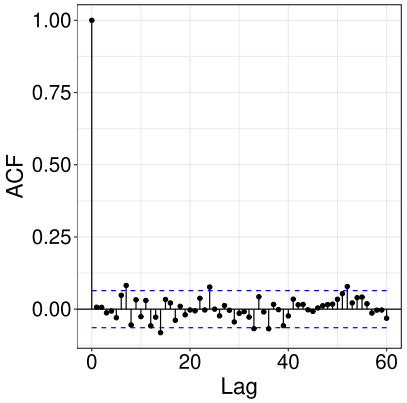
1. The increase in one unit ($\mu g/m^3$) of weekly average of PM10 increases, in average, 0.2 the number of weekly hospitalizations.
2. The increase in one unit ($\mu g/m^3$) of weekly minimum of NO_2 increases, in average, 0.81 the number of weekly hospitalizations.
3. The increase in one unit ($\mu g/m^3$) of weekly maximum of NO_2 increases, in average, 0.06 the number of weekly hospitalizations.
4. The increase in one unit ($\mu g/m^3$) of weekly average of NO decreases, in average, 0.21 the number of weekly hospitalizations.



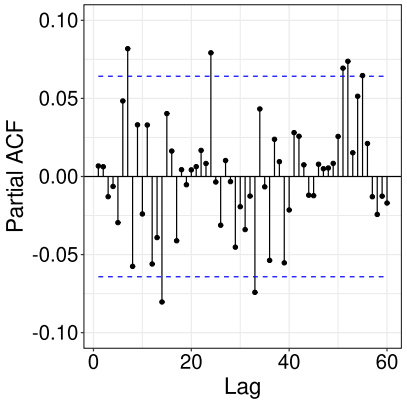
(a) index plot



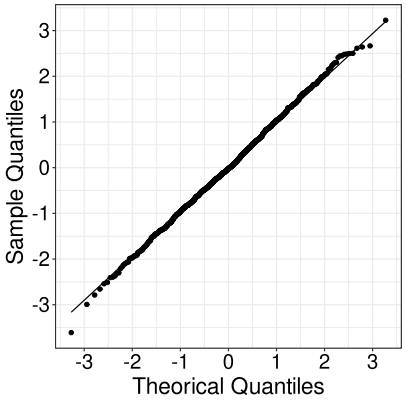
(b) histogram and estimated density



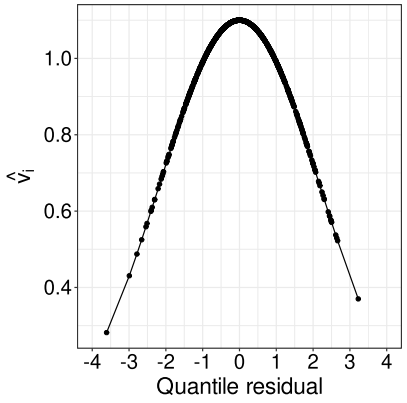
(c) autocorrelation function



(d) partial autocorrelation function



(e) normal probability plot



(f) $\hat{v}_i$ weight vs quantile residual

Fig. 2 Quantile residual analysis of the model (9) fitted to the hospitalization data

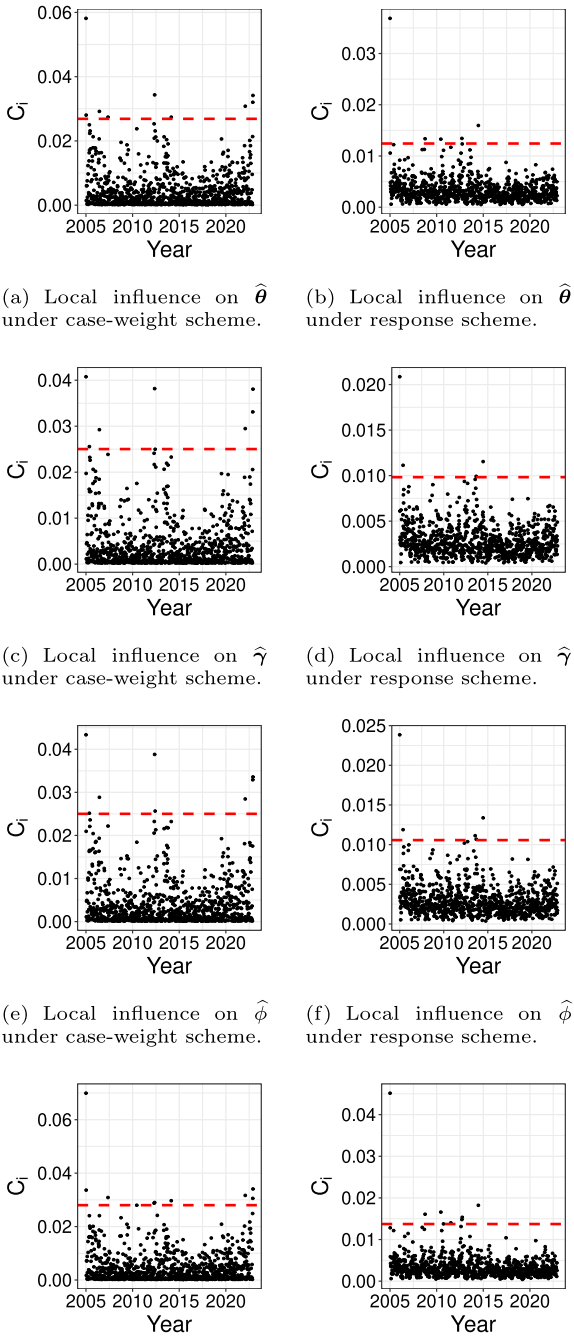


Fig. 3 Index plots of C_i for assessing the local influence on the MPLEs under case-weight and response perturbation schemes in the model (9) fitted to the hospitalization data

(a) Local influence on $\widehat{\boldsymbol{\theta}}$ under case-weight scheme.

(b) Local influence on $\widehat{\boldsymbol{\theta}}$ under response scheme.

(c) Local influence on $\widehat{\boldsymbol{\gamma}}$ under case-weight scheme.

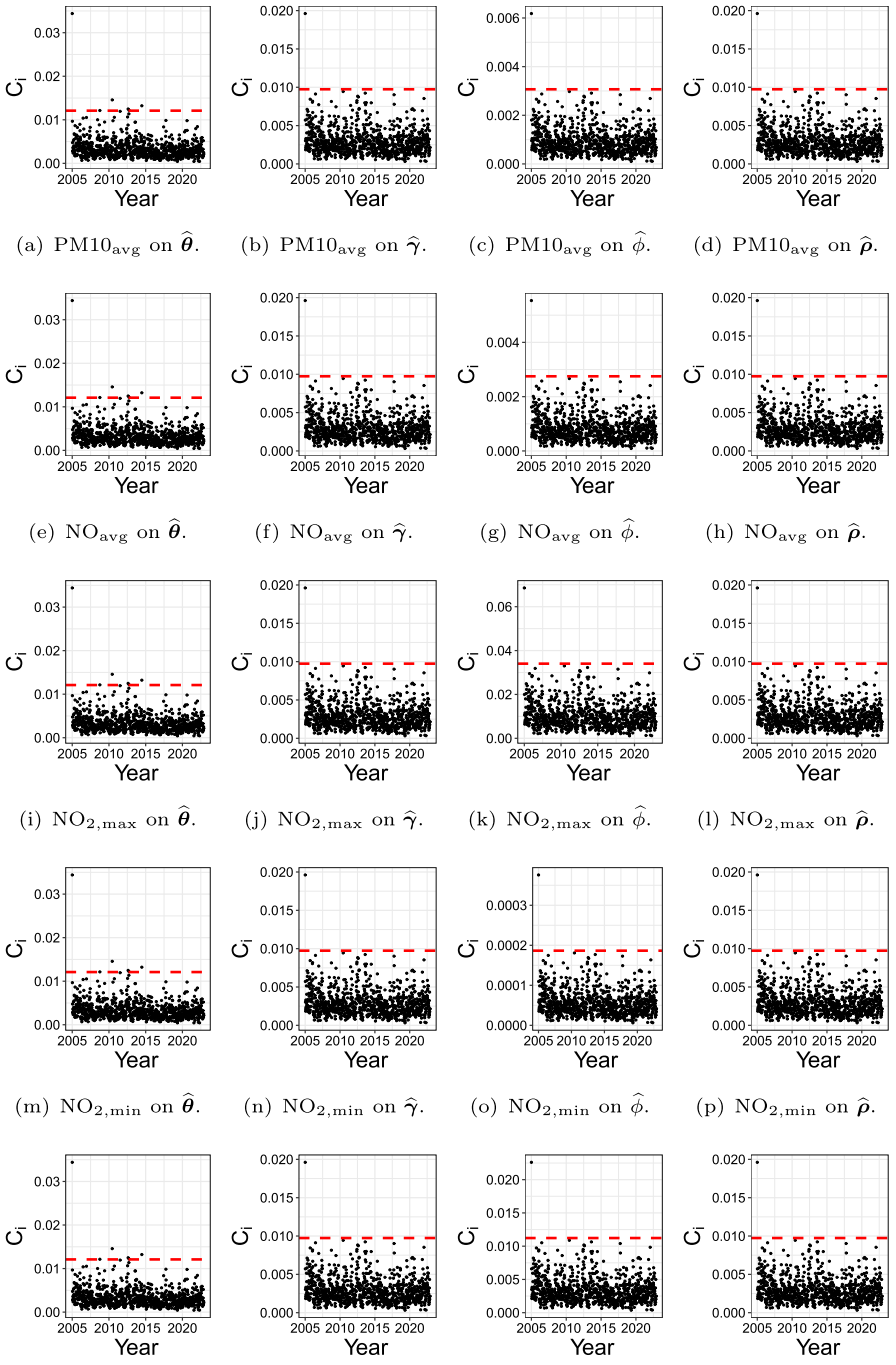
(d) Local influence on $\widehat{\boldsymbol{\gamma}}$ under response scheme.

(e) Local influence on $\widehat{\phi}$ under case-weight scheme.

(f) Local influence on $\widehat{\phi}$ under response scheme.

(g) Local influence on $\widehat{\boldsymbol{\rho}}$ under case-weight scheme.

(h) Local influence on $\widehat{\boldsymbol{\rho}}$ under response scheme.



(a) $\mathrm{PM10_{avg}}$ on $\widehat{\boldsymbol{\theta}}$. (b) $\mathrm{PM10_{avg}}$ on $\widehat{\boldsymbol{\gamma}}$. (c) $\mathrm{PM10_{avg}}$ on $\widehat{\phi}$. (d) $\mathrm{PM10_{avg}}$ on $\widehat{\boldsymbol{\rho}}$.

(e) $\mathrm{NO_{avg}}$ on $\widehat{\boldsymbol{\theta}}$. (f) $\mathrm{NO_{avg}}$ on $\widehat{\boldsymbol{\gamma}}$. (g) $\mathrm{NO_{avg}}$ on $\widehat{\phi}$. (h) $\mathrm{NO_{avg}}$ on $\widehat{\boldsymbol{\rho}}$.

(i) $\mathrm{NO_{2,max}}$ on $\widehat{\boldsymbol{\theta}}$. (j) $\mathrm{NO_{2,max}}$ on $\widehat{\boldsymbol{\gamma}}$. (k) $\mathrm{NO_{2,max}}$ on $\widehat{\phi}$. (l) $\mathrm{NO_{2,max}}$ on $\widehat{\boldsymbol{\rho}}$.

(m) $\mathrm{NO_{2,min}}$ on $\widehat{\boldsymbol{\theta}}$. (n) $\mathrm{NO_{2,min}}$ on $\widehat{\boldsymbol{\gamma}}$. (o) $\mathrm{NO_{2,min}}$ on $\widehat{\phi}$. (p) $\mathrm{NO_{2,min}}$ on $\widehat{\boldsymbol{\rho}}$.

(q) RH on $\widehat{\boldsymbol{\theta}}$. (r) RH on $\widehat{\boldsymbol{\gamma}}$. (s) RH on $\widehat{\phi}$. (t) RH on $\widehat{\boldsymbol{\rho}}$.

Fig. 4 Index plots of C_i for assessing the local influence on the MPLEs under explanatory variable perturbations in the model (9) fitted to the hospitalization data

Table 3 Parameter estimates (parametric part) and their approximate standard errors from the model (9) fitted to the hospitalization data

Covariate	Estimate	Std. Error	z-value	P-value
Intercept	67.5151	3.8899	17.3566	< 0.0001
$PM10_{avg}$	0.1998	0.0629	3.1787	0.0015
$NO_{2,min}$	0.8105	0.2160	3.7532	0.0002
$NO_{2,max}$	0.0599	0.0257	2.3341	0.0198
NO_{avg}	-0.2112	0.0871	-2.4236	0.0156
RH_{min}	0.1044	0.0513	2.0341	0.0422
Parameter				
ρ_1	0.4270	0.0326	13.1084	< 0.0001
ρ_2	0.2010	0.0350	5.7348	< 0.0001
ρ_3	0.1234	0.0355	3.4806	0.0005
ρ_4	0.0893	0.0327	2.7306	0.0064
ϕ	128.7017	6.7977	18.9331	< 0.0001

5. The increase in one unit (1%) of weekly minimum RH increases, in average, 0.10 the number of weekly hospitalizations.

We are assuming for each interpretation above that the remaining effects are fixed. So, in accordance with the findings from previous studies (Negrisoli and Nascimento 2013; Schwartz 2004; Amâncio and Costa Nascimento 2012), we find a positive association between short and medium-term exposure to average of PM10, minimum of NO_2 and maximum of NO_2 with respiratory disease hospitalizations. The fact that maximum of NO_2 presents a smaller positive effect than minimum of NO_2 could be explained by a potential saturation effect by smaller effects of changes in weekly maximum of NO_2 in a city with high baseline levels of NO_2, as Sorocaba is classified due to its high industrialization. Several epidemiological studies demonstrate the harmful effects of atmospheric pollutants on human health in terms of aggravation of respiratory diseases, in addition to their seasonal behavior, with higher concentrations in the winter months, when air quality is more compromised due to the impaired dispersion of pollutants consequence of rain absence, lower relative humidity and wind. In Sorocaba, the month with the highest relative humidity is February (79.29%) and the month with the lowest relative humidity is August (64.72%), whereas the month with the highest number of rainy days is January (20.67 days) and the month with the smallest number is August (4.60 days).[1]

Additionally, Table 4 shows that both smooth functions are significant and Fig. 5 describes the 95% pointwise confidence bands for trend and seasonality. We observe a significant seasonality with higher hospitalization from week 20 to week 30 which agrees with the winter season described above. Regarding to the trend component, controlling the seasonality the trend increases significantly after 2020. We may conjecture two possibilities: (1) confounding factor of COVID19 and (2) because pollution has steadily risen in the past years and the covariates included into the model are not sufficient to explain all effects on hospitalization.

[1] https://pt.climate-data.org/america-do-sul/brasil/sao-paulo/sorocaba-756/.

Table 4 Wald statistics, respective p-values and effective degrees of freedom of the additive components from the model (9) fitted to the hospitalization data

	Wald	$df_s(\lambda)$	p-value
f_1	22.5479	4.0305	< 0.0001
f_2	36.1920	2.1655	< 0.0001

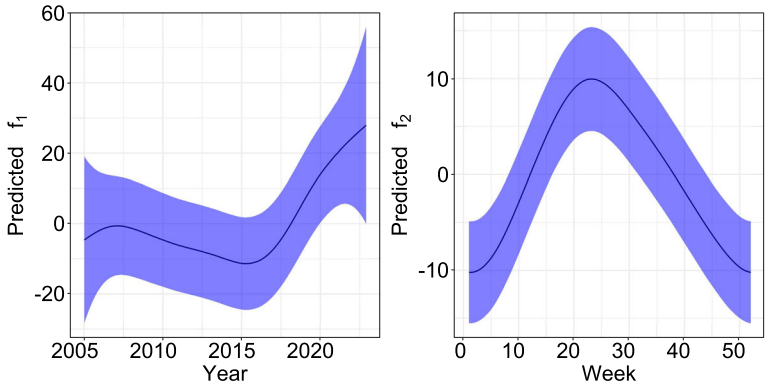


Fig. 5 Pointwise confidence bands of 95% for trend (left) and seasonality (right) components from the model (9) fitted to the hospitalization data

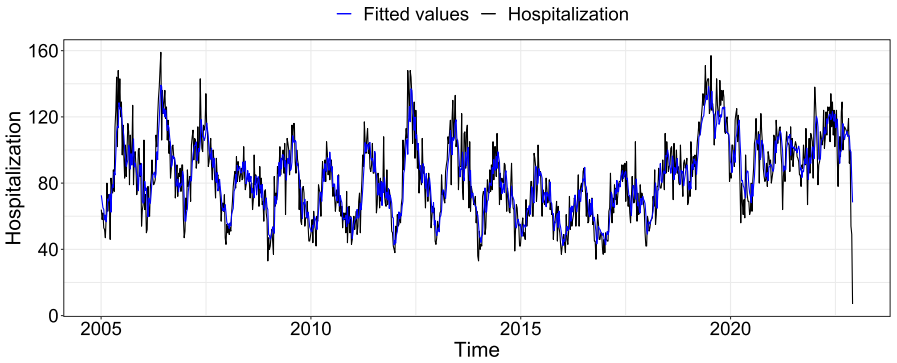


Fig. 6 Observed (black) and estimated (blue) hospitalization from the model (9) fitted to the hospitalization data

Finally, Fig. 6 describes the observed hospitalization time series and the estimated hospitalization from the fitted model and we may observe a very close agreement between them.

7 Concluding remarks

The additive model with symmetric autoregressive errors in which trend and seasonality of time series are decomposed, proposed by Oliveira and Paula (2021), is extended in this paper by including a linear component and general additive terms. A novel

iterative process that combines the P-GAM algorithm proposed by Marx and Eilers (1998) with a quasi-Newton algorithm is derived. Inferential procedures, residual analysis and sensitivity studies are given as well as various additional results are given as n material. The weekly hospitalization for respiratory diseases in Sorocaba city, São Paulo, Brazil, is modeled as an application using climate and pollution explanatory variables. It is possible to find, from the fitted model, an association between the mean response with exposure to the atmospheric pollutants $PM10_{avg}$, $NO_{2,min}$, $NO_{2,max}$, NO_{avg} and RH_{min}, and hospital admissions for respiratory diseases, may provide useful information for the development of policies to reduce public health risks. The authors are developing a R package with the procedures presented in this paper and possible future works include forecast procedures and extensions for single-index and spatial models.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00362-024-01590-w.

Funding The authors did not receive support from any organization for the submitted work.

Declarations

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Amâncio CT, Costa Nascimento LF (2012) Asthma and air pollutants: a time series study. Revista da Associação Médica Brasileira (English Edition) 58(3):302–307

Barros M, Paula GA (2019) Discussion of "Birnbaum-Saunders distributions: a review of models, analysis and applications". Appl Stoch Model Bus Ind 35(1):96–99

Basu R (2009) High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. Environ Health 8(1):40

Basu R, Samet JM (2002) 12. Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. Epidemiol Rev 24(2):190–202

Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong BG (2013) Time series regression studies in environmental epidemiology. Int J Epidemiol 42:1187–1195

Byrd RH, Lu P, Nocedal J, Zhu C (1995) A limited memory algorithm for bound constrained optimization. SIAM J Sci Comput 16(5):1190–1208

Cardozo CA, Paula GA, Vanegas LH (2022) Generalized log-gamma additive partial linear models with p-spline smoothing. Stat Pap 63(6):1953–1978

Cheng S, Chen J, Liu X (2019) GMM estimation of partially linear single-index spatial autoregressive model. Spat Stat 31:100354

Cysneiros FJA, Paula GA (2005) Restricted methods in symmetrical linear regression models. Comput Stat Data Anal 49(3):689–708

Davidon WC (1991) Variable metric method for minimization. SIAM J Optim 1(1):1–17

Dunn PK, Smyth GK (1996) Randomized quantile residuals. J Comput Graph Stat 5(3):236–244

Eilers PH, Marx BD (1996) Flexible smoothing with b-splines and penalties. Stat Sci 11(2):89–102

Fang KT, Anderson TW (1990) Statistical inference in elliptically contoured and related distributions. Allerton Press, New York

Green PJ, Silverman BW (1994) Nonparametric regression and generalized linear models: a roughness penalty approach. Chapman and Hall/CRC, London

Hastie T, Tibshirani R (1990) Generalized additive models. Chapman & Hall, London

Huang L, Xia Y, Qin X (2016) Estimation of semivarying coefficient time series models with ARMA errors. Ann Stat 44(4):1618–1660
Huang L, Jiang H, Wang H (2019) A novel partial-linear single-index model for time series data. Comput Stat Data Anal 134:110–122
Lancaster P, Salkauskas K (1986) Curve and surface fitting. An introduction. Academic Press, London
Lee SY, Xu L (2004) Influence analyses of nonlinear mixed-effects models. Comput Stat Data Anal 45(2):321–341
Lim JT, Tan KB, Abisheganaden J, Dickens BL (2023) Forecasting upper respiratory tract infection burden using high-dimensional time series data and forecast combinations. PLoS Comput Biol 19(2):1–14
Liu S (2000) On local influence for elliptical linear models. Stat Pap 41(2):211–224
Liu S (2004) On diagnostics in conditionally heteroskedastic time series models under elliptical distributions. J Appl Probab 41(A):393–405
Marx BD, Eilers PH (1998) Direct generalized additive modeling with penalized likelihood. Comput Stat Data Anal 28(2):193–209
Mittelhammer RC, Judge GG, Miller DJ (2000) Econometric foundations. Cambridge University Press, Cambridge
Negrisoli J, Nascimento LFC (2013) Atmospheric pollutants and hospital admissions due to pneumonia in children. Revista Paulista de Pediatria 31:501–506
Oliveira RA, Paula GA (2021) Additive models with autoregressive symmetric errors based on penalized regression splines. Comput Stat 36(4):2435–2466
Poon WY, Poon YS (1999) Conformal normal curvature and assessment of local influence. J R Stat Soc Ser B (Stat Methodol) 61(1):51–61
Relvas CEM, Paula GA (2016) Partially linear models with first-order autoregressive symmetric errors. Stat Pap 57(3):795–825
Ruppert D, Wand MP, Carroll RJ (2003) Semiparametric regression. Number 12. Cambridge University Press
Schwartz J (2004) Air pollution and children's health. Pediatrics 113(4 Suppl):1037–1043
Vanegas LH, Paula GA (2016) An extension of log-symmetric regression models: R codes and applications. J Stat Comput Simul 86(9):1709–1735
Wood SN (2017) Generalized additive models: an introduction with R, 2nd edn. Chapman and Hall/CRC, London
Ye X, Wolff R, Yu W, Vaneckova P, Pan X, Tong S (2012) Ambient temperature and morbidity: a review of epidemiological evidence. Environ Health Perspect 120(1):19–28
Yu Y, Ruppert D (2002) Penalized spline estimation for partially linear single-index models. J Am Stat Assoc 97(460):1042–1054

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.