Computational Statistics
https://doi.org/10.1007/s00180-021-01130-2

ORIGINAL PAPER

Estimation and diagnostic for partially linear models with first-order autoregressive skew-normal errors

Clécio da Silva Ferreira[1] · **Gilberto A. Paula**[2] · **Gustavo C. Lana**[1]

Received: 18 June 2020 / Accepted: 29 June 2021

Abstract
Estimation and diagnostic procedures for partially linear models with first-order autoregressive [AR(1)] skew-normal errors are proposed in this paper. An EM iterative process with analytic expressions for the M and E-steps, which combines back-fitting and Newton–Raphson algorithms, is developed for the parameter estimation. A linear smoother for the estimation of the effective degrees of freedom concerning the non-parametric component is derived from the iterative process. Local influence analysis is developed based on the conditional expectation of the complete-data log-likelihood function, used in the EM algorithm. A simulation study is also conducted to evaluate the efficiency of the EM algorithm. Finally, the methodology developed through the paper is illustrated with a real data set on daily ozone concentration.

Keywords Back-fitting algorithm · EM-algorithm · Local influence · Penalized Smoothing · Semiparametric models · Skew-normal distribution

1 Introduction

The aim of this paper is to develop an Expectation–Maximization (EM) iterative process and local influence procedures for partially linear models (PLMs) (or semi-parametric models) with first-order autoregressive [AR(1)] skew-normal errors.

Linear and nonlinear models with AR errors have been studied by various authors. For example, under elliptical distributions, Liu (2004) developed diagnostic methods in

---

✉ Clécio da Silva Ferreira
clecio.ferreira@ufjf.edu.br

Gilberto A. Paula
giapaula@ime.usp.br

Gustavo C. Lana
gustavo.lana@ufjf.edu.br

[1] Department of Statistics, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil

[2] Institute of Mathematics and Statistics, University of São Paulo, São Paulo, SP, Brazil

conditional heteroscedastic time series models, Paula et al. (2009) discussed estimation and diagnostics in linear models with AR(1) errors, whereas Cao et al. (2010) presented procedures for assessing heteroscedasticity and autocorrelation as well as diagnostic methods in nonlinear models with AR(1) errors. In the asymmetric context, Bondon (2009) introduced an autoregressive model with epsilon-skew-normal innovations, whereas Sharafi and Nematollahi (2016) considered an autoregressive model of order one with skew-normal innovations.

On the other hand, regression models with AR(1) errors may present a nonlinear relationship between the response and some continuous explanatory variables, such as time. So, nonparametric forms may be added into the systematic component of the model in order to accommodate this covariate flexibly. Such models, named partially linear models with AR errors, have been investigated by various authors. For example, Ferreira et al. (2013) developed a Bayesian analysis for partially linear models with first-order autoregressive errors belonging to the class of the scale mixtures of normal distributions whereas Relvas and Paula (2016) proposed an iterative process and some diagnostic procedures for partially linear models with AR(1) symmetric errors. In the asymmetric context, Ferreira and Paula (2017) developed a PLM under a skew-normal distribution allowing maximum likelihood estimation via EM algorithm and diagnostic analysis using Zhu and Lee's approach (Zhu and Lee 2001).

In this paper, we propose a generalization of the PLMs with AR(1) errors under a skew-normal distribution (Sahu et al. 2003) using the EM algorithm for the parameter estimation and the local influence approach developed by Zhu and Lee (2001). As pointed out by Greene (2012) AR(1) error models have been largely applied in time series analysis due to their flexibility of modeling more complex processes. The remainder of the paper is structured as follows. Section 2 presents the proposed partially linear models and the penalized log-likelihood function with AR(1) skew-normal errors and some related properties. An EM iterative process for the parameter estimation is derived in Section 3, combining back-fitting and Newton–Raphson algorithms for estimating the nonparametric and parametric components. A linear smoother for the estimation of the effective degrees of freedom of the nonparametric component is derived from the EM iterative process. In Sect. 4, we derive local influence curvatures using the methodology proposed by Zhu and Lee (2001) under some usual perturbation schemes. Section 5 deals with a simulation study for evaluating the efficiency of the EM algorithm and Sect. 6 illustrates the application of the proposed methodology in real data on daily ozone concentration. Section 7 summarizes the contributions of the paper.

2 Specification of the model

The skew-normal (SN) distribution (Azzalini 1985) has range for asymmetry varying between −0.995 and 0.995 and the maximum of the kurtosis is 3.869. So, this distribution is useful to modeling data with asymmetry and some degree of kurtosis. There are some other versions of this distribution (see Nadarajah and Kotz 2003; Genton 2004; Ferreira and Steel 2006), among others. In this work we use another version of the skew-normal, developed by Sahu et al. (2003), where there is a one-to-one relationship between the parameters of the two versions to the univariate case. An advantage of this version is that it directly allows analytic expressions in the M and E-steps for the EM algorithm in the proposed model.

A random variable Y follows a univariate skew-normal distribution (Sahu et al. 2003) with location parameter μ, scale parameter σ^2 and skewness parameter δ, if its probability density function (pdf) takes the form

$$f(y|\mu,\sigma^2,\delta)=\frac{2}{\sqrt{\sigma^2+\delta^2}}\phi\left(\frac{y-\mu}{\sqrt{\sigma^2+\delta^2}}\right)\Phi\left(\frac{\delta}{\sigma}\frac{y-\mu}{\sqrt{\sigma^2+\delta^2}}\right), \tag{1}$$

where ϕ and Φ are, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of the $N(0,1)$, namely $Y\sim SN(\mu,\sigma^2,\delta)$. For $\delta=0$, the pdf in (1) reduces to the one of the normal distribution. The mean and the variance of Y are, respectively, given by

$$E(Y)=\mu+b\delta \quad \text{and} \quad Var(Y)=\sigma^2+(1-b^2)\delta^2, \tag{2}$$

where $b=\sqrt{2/\pi}$. The stochastic representation is given by $Y\overset{d}{=}\mu+\delta|X_0|+\sigma X_1$, where X_0 and X_1 are independent random variables $N(0,1)$. The cumulative density function of $Y\sim SN(\mu,\sigma^2,\delta)$ is given by

$$F_Y(y;\mu,\sigma^2,\delta)=2\Phi_2\left(\left(\frac{y-\mu}{\sqrt{\sigma^2+\delta^2}},0\right);\mathbf{0},\mathbf{\Omega}\right), \tag{3}$$

where $\mathbf{\Omega}=\begin{bmatrix}1 & -\delta_1\\ -\delta_1 & 1\end{bmatrix}$, with $\delta_1=\frac{\delta}{\sqrt{\sigma^2+\delta^2}}$.

The partially linear model with AR(1) skew-normal errors is defined as follows:

$$\begin{aligned} y_i &= \mu_i+\epsilon_i,\\ \epsilon_i &= \rho\epsilon_{i-1}+e_i, \quad -1<\rho<1, \quad \text{and}\\ e_i &\sim SN(-b\delta,\sigma^2,\delta), \quad i=1,\ldots,n, \end{aligned} \tag{4}$$

where y_i's denote the response values, $\mu_i=\mathbf{x}_i^\top\boldsymbol{\beta}+f(t_i)$ is the expected value of Y_i, $\mathbf{x}_i=(x_{i1},\ldots,x_{ip})^\top$ contains values of explanatory variables, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)^\top$ is the fixed-parameter vector, ρ is the AR coefficient, t_i's represent values of a continuous covariate that has some nonlinear relationship with the response, for example, the time, $f(\cdot)$ denotes a smoothing function and e_i's are independent random errors with a skew-normal distribution of zero mean, for $i=1,\ldots,n$. We will assume $\epsilon_0=0$ and when $\rho=0$ model (4) reduces to the skew-normal partially linear model discussed in Ferreira and Paula (2017).

Properties:

1. $Y_1\sim SN(\mu_1-b\delta,\sigma^2,\delta)$. So, we have that $E(Y_1)=\mu_1$ and $Var(Y_1)=\sigma^2+(1-b^2)\delta^2$.
2. $Y_i|y_{i-1}\sim SN(\mu_i+\rho(y_{i-1}-\mu_{i-1})-b\delta,\sigma^2,\delta), \quad i=2,\ldots,n$.
3. $E(Y_i)=\mu_i$ and $Var(Y_i)=Var(Y_1)\sum_{j=1}^{i}\rho^{2(i-j)}$, for $i=2,\ldots,n$.

4. For $k > i$, $Cov(Y_i, Y_k) = Var(Y_1)\sum_{j=0}^{i} \rho^{i+k-2j}$.

This paper considers smoothing functions expressed as B-splines (de Boor 2001) and (Wood 2017), namely $f(t) = \sum_{j=1}^{q} N_{j,k}(t)\gamma_j$, where

$$N_{j,1}(t) = \begin{cases} 1, & \kappa_j \le t < \kappa_{j+1}, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$N_{j,k}(t) = w_{jk}N_{j,k-1}(t) + (1 - w_{j+1,k})N_{j+1,k-1}(t), \quad \text{for} \quad k > 1,$$

with $w_{jk}(t) = \frac{t-\kappa_j}{\kappa_{j+k-1}-\kappa_j}$, $N_{j,k}(t)$ denoting the B-spline basis functions of degree k and γ_j are coefficients to be estimated, for $j = 1, \ldots, q$, whereas m is the number of internal knots, namely $a_1 \le \kappa_1 < \cdots < \kappa_m \le b_1$ and $q = m - k - 1$. For simplicity, we will assume $k = 3$ (cubic B-splines) and consequently, $f(t) = \sum_{j=1}^{q} N_j(t)\gamma_j$ with $N_j(t) = N_{j,3}(t)$. B-splines have been largely applied for modeling nonlinear relationships due to their flexibility and local control (Eilers and Marx 1996; Wood 2017).

Then, the systematic component of the model (3) may be expressed in the matrix form $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\boldsymbol{\gamma}$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, $\mathbf{X}$ denotes an $n \times p$ matrix with rows $\mathbf{x}_i^\top$, $\mathbf{N}$ is an $n \times q$ matrix with rows $\mathbf{n}_i^\top = (N_1(t_i), \ldots, N_q(t_i))$, for $i = 1, \ldots, n$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$. In order to obtain a smoothing fitting, we will consider a continuous penalization in the likelihood function from which an EM iterative process will be developed. However, discrete penalization has been largely applied with B-splines, such as P-splines (Eilers and Marx 1996). See Appendix B for the construction of the matrix $\mathbf{N}$ based on the equidistant knots $\boldsymbol{\kappa}$ and values $\mathbf{t} = (t_1, \ldots, t_n)$.

3 Parameter estimation via EM-algorithm

From (4) the observed-data log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \sigma^2, \delta, \rho)^\top \in \mathbb{R}^{p^*}$, $p^* = p + q + 3$ may be expressed as

$$\ell(\boldsymbol{\theta}) = n \log \frac{2}{\sqrt{2\pi}} - \frac{n}{2}\log(\sigma^2 + \delta^2) - \frac{\sigma^2}{2\delta^2}\sum_{i=1}^{n} B_i^2 + \sum_{i=1}^{n} \log \Phi(B_i), \tag{5}$$

where $B_i = \frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}(y_i - \xi_i + b\delta)$, $\xi_i = \mu_i + \rho(y_{i-1} - \mu_{i-1})$, $\mu_i = \mathbf{x}_i^\top\boldsymbol{\beta} + \mathbf{n}_i^\top\boldsymbol{\gamma}$ with $\mathbf{x}_i^\top$ and $\mathbf{n}_i^\top$ being the i-th row of $\mathbf{X}$ and $\mathbf{N}$, respectively, for $i = 1, \ldots, n$.

One may have two difficulties in maximizing $\ell(\boldsymbol{\theta})$ in (4), the evaluation of the term $\Phi(\cdot)$ and the need of imposing some restriction for $f(t)$ to avoid over-fitting and nonidentification of $\boldsymbol{\gamma}$ (see, for instance, Green 1987). So, we will propose the parameter estimation evaluated by an EM-algorithm with the incorporation into the log-likelihood function of a penalty function $J(\boldsymbol{\gamma})$ together with a smoothing parameter $\alpha > 0$.

3.1 EM-algorithm

The EM algorithm (Dempster et al. 1977) has been largely applied for finding maximum likelihood estimates. One advantage of the EM algorithm is that the M-step involves only complete data ML estimation, which is often computationally simple.

Let $TN(\mu, \sigma^2; 0, +\infty)$ denote the truncated-normal distribution with parameters μ and σ^2 and support in $(0, +\infty)$ (Johnson et al. 1994). Using the stochastic representation of skew-normal distribution and Property 2, we have that

$$\begin{aligned} Y_i|Z_i &= z_i, y_{i-1} \overset{ind}{\sim} N(\xi_i - b\delta + \delta z_i, \sigma^2) \quad \text{and} \\ Z_i &\overset{iid}{\sim} TN(0, 1; (0, +\infty)), \quad i = 1, \ldots, n. \end{aligned}$$

So, after some algebraic manipulations, the joint distribution for (Y_i, Z_i) becomes given by

$$\begin{aligned} f(y_i, z_i; y_{i-1}, \boldsymbol{\theta}) &= 2\phi(y_i|\xi_i + \delta z_i - b\delta, \sigma^2)\phi(z_i)\mathbb{I}_{(0,+\infty)}(z_i) \\ &= 2\phi(y_i|\xi_i - b\delta, \sigma^2 + \delta^2)\phi\left(z_i|\mu_{iz}, \sigma_z^2\right)\mathbb{I}_{(0,+\infty)}(z_i), \end{aligned}$$

where $\mu_{iz} = \frac{\delta}{\sigma^2+\delta^2}(y_i - \xi_i + b\delta)$ and $\sigma_z^2 = \frac{\sigma^2}{\sigma^2+\delta^2}$. So, we have that $Z_i|y_i, \boldsymbol{\theta} \sim TN\left(\mu_{iz}, \sigma_z^2; (0, +\infty)\right)$. Thus, from properties of truncated normal distributions, we obtain

$$\begin{aligned} E[Z_i|y_i] &= \mu_{iz} + \sigma_z W_\Phi(\mu_{iz}/\sigma_z) \quad \text{and} \\ E[Z_i^2|y_i] &= \mu_{iz}^2 + \sigma_z^2 + \sigma_z\mu_{iz}W_\Phi(\mu_{iz}/\sigma_z), \end{aligned} \tag{6}$$

where $W_\Phi(u) = \phi(u)/\Phi(u)$.

Let $\mathbf{y} = (y_1, \ldots, y_n)^\top$ and $\mathbf{z} = (z_1, \ldots, z_n)^\top$ with $\mathbf{z}$ being treated as missing data. Then, the complete log-likelihood function associated with $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{z}^\top)^\top$ may be expressed as

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = C - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(y_i - \xi_i)^2 - 2\delta(y_i - \xi_i)(z_i - b) + \delta^2(b^2 - 2bz_i + z_i^2)\right], \tag{7}$$

where C is a constant that does not depend on unknown parameters.

In the E-step, one has the function

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}] \propto -\frac{n}{2}\log\sigma^2 \\ -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(y_i - \xi_i)^2 - 2\delta(y_i - \xi_i)(\widehat{z}_i^{(k)} - b) + \delta^2(b^2 - 2b\widehat{z}_i^{(k)} + \widehat{z^2}_i^{(k)})\right], \end{aligned} \tag{8}$$

where $\widehat{z}_i^{(k)} = \mathrm{E}[Z_i|y_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ and $\widehat{z^2}_i^{(k)} = \mathrm{E}[Z_i^2|y_i, \widehat{\boldsymbol{\theta}}^{(k)}]$.

Similarly to Green (1990), the maximum penalized likelihood estimate (MPLE) of $\boldsymbol{\theta}$ will be obtained by maximizing the function

$$Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) - \frac{\alpha}{2}J(\boldsymbol{\gamma}), \tag{9}$$

where $\alpha > 0$ denotes a smoothing parameter whereas $J(\boldsymbol{\gamma})$ denotes the penalty function, which is defined as follows

$$J(\boldsymbol{\gamma}) = \int_{a_1}^{b_1} [f^{(2)}(t)]^2 dt = \boldsymbol{\gamma}^\top \mathbf{K}\boldsymbol{\gamma},$$

(see also Wood 2017, Chap. 5) with $\mathbf{K} \in \mathbb{R}^{q\times q}$ being a nonnegative definite matrix that depends only on the knot differences (see Appendix B for details). Using (6), we have that

$$\widehat{z}_i^{(k)} = \widehat{\mu}_{iz}^{(k)} + \widehat{\sigma}_z^{(k)} W_\Phi\left(\frac{\widehat{\mu}_{iz}^{(k)}}{\widehat{\sigma}_z^{(k)}}\right) \tag{10}$$

and

$$\widehat{z^2}_i^{(k)} = [\widehat{\mu}_{iz}^{(k)}]^2 + \widehat{\sigma^2}_z^{(k)} + \widehat{\sigma}_z^{(k)}\widehat{\mu}_{iz}^{(k)} W_\Phi\left(\frac{\widehat{\mu}_{iz}^{(k)}}{\widehat{\sigma}_z^{(k)}}\right), \tag{11}$$

where $\widehat{\mu}_{iz}^{(k)} = \frac{\widehat{\delta}^{(k)}}{\widehat{\sigma^2}^{(k)}+\widehat{\delta}^{(k)2}}(y_i - \widehat{\xi}_i^{(k)} + b\widehat{\delta}^{(k)})$, $\widehat{\sigma^2}_z^{(k)} = \frac{\widehat{\sigma^2}^{(k)}}{\widehat{\sigma^2}^{(k)}+\widehat{\delta}^{(k)2}}$ and $\widehat{\xi}_i^{(k)} = \widehat{\mu}_i^{(k)} + \widehat{\rho}^{(k)}(y_{i-1} - \widehat{\mu}_{i-1}^{(k)})$, with $\widehat{\mu}_i^{(k)} = \mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}^{(k)} + \mathbf{n}_i^\top\widehat{\boldsymbol{\gamma}}^{(k)}$, $i = 1, \ldots, n$.

Let $\mathbf{A} = \mathbf{A}(\rho)$ an $(n \times n)$ matrix given by

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

The M-step maximizes $Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$, obtaining a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as described below:

E-step: Compute, for $i = 1, \ldots, n$, $\widehat{z}_i^{(k)}$ and $\widehat{z^2}_i^{(k)}$ using (10)–(11).

M-step: Update $\widehat{\boldsymbol{\beta}}^{(k)}$, $\widehat{\boldsymbol{\gamma}}^{(k)}$, $\widehat{\sigma^2}^{(k)}$, $\widehat{\delta}^{(k)}$ and $\widehat{\rho}^{(k)}$, respectively, as

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \widehat{\mathbf{S}}_\beta^{(k)}(\alpha)\left[\widehat{\mathbf{y}}^{*(k)} - \widehat{\mathbf{A}}^{(k)}\mathbf{N}\widehat{\boldsymbol{\gamma}}^{(k)}\right],$$

$$\widehat{\boldsymbol{\gamma}}^{(k+1)} = \widehat{\mathbf{S}}_f^{(k)}(\alpha)\left[\widehat{\mathbf{y}}^{*(k)} - \widehat{\mathbf{A}}^{(k)}\mathbf{X}\widehat{\boldsymbol{\beta}}^{(k)}\right],$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \Big[(y_i - \widehat{\xi}_i^{(k)})^2 - 2\widehat{\delta}^{(k)}(y_i - \widehat{\xi}_i^{(k)})(\widehat{z}_i^{(k)} - b)$$

$$
\begin{aligned}
&\quad + \widehat{\delta}^{(k)^2}\left(b^2 - 2b\widehat{z}_i^{(k)} + \widehat{z^2}_i^{(k)}\right)\Big],\\
\widehat{\delta}^{(k+1)} &= \frac{\sum_{i=1}^{n}(y_i - \widehat{\xi}_i^{(k)})(\widehat{z}_i^{(k)} - b)}{\sum_{i=1}^{n}(b^2 - 2b\widehat{z}_i^{(k)} + \widehat{z^2}_i^{(k)})} \quad \text{and}\\
\widehat{\rho}^{(k+1)} &= \frac{\sum_{i=1}^{n}[r_i - \widehat{\delta}^{(k)}(\widehat{z}_i^{(k)} - b)]r_{i-1}}{\sum_{i=1}^{n} r_{i-1}^2},
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\mathbf{S}}_{\beta}^{(k)}(\alpha) &= \left[(\widehat{\mathbf{A}}^{(k)}\mathbf{X})^\top(\widehat{\mathbf{A}}^{(k)}\mathbf{X})\right]^{-1}(\widehat{\mathbf{A}}^{(k)}\mathbf{X})^\top \quad \text{and}\\
\widehat{\mathbf{S}}_{\boldsymbol{\gamma}}^{(k)}(\alpha) &= \left[(\widehat{\mathbf{A}}^{(k)}\mathbf{N})^\top(\widehat{\mathbf{A}}^{(k)}\mathbf{N}) + \alpha\widehat{\sigma^2}^{(k)}\mathbf{K}\right]^{-1}(\widehat{\mathbf{A}}^{(k)}\mathbf{N})^\top
\end{aligned}
$$

are smoother with $\widehat{\mathbf{y}}^{*(k)} = \widehat{\mathbf{A}}^{(k)}\mathbf{y} - \widehat{\delta}^{(k)}(\widehat{\mathbf{z}}^{(k)} - b)$ being a pseudo-response, $r_i = y_i - \widehat{\mu}_i^{(k)} = y_i - \mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}^{(k)} - \mathbf{n}_i^\top\widehat{\boldsymbol{\gamma}}^{(k)}, i = 1, \ldots, n, r_0 = 0$, and $\widehat{\mathbf{z}}^{(k)} = (\widehat{z}_1^{(k)}, \ldots, \widehat{z}_n^{(k)})^\top$, for $k = 0, 1, 2, \ldots$.

Note that, at the **M-step**, the iterative process above combines a back-fitting algorithm for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ with a Newton–Raphson algorithm for estimating σ^2, δ and ρ. The iterations are repeated until a suitable convergence rule is satisfied, e.g., $||\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}||$ is sufficiently small, say 10^{-6}. A set of reasonable starting values may be achieved by computing $\widehat{\boldsymbol{\beta}}^{(0)}$ and $\widehat{\sigma}^{2(0)}$ as the solution of the least-squares regression model of $\mathbf{y}$ on $\mathbf{X}$. So, for fixed α, $\widehat{\boldsymbol{\gamma}}^{(0)} = (\mathbf{N}^\top\mathbf{N} + \alpha\widehat{\sigma}^{2(0)}\mathbf{K})^{-1}\mathbf{N}^\top(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(0)})$, $\widehat{\delta}^{(0)}$ may be the sample skewness coefficient of $\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(0)} - \mathbf{N}\widehat{\boldsymbol{\gamma}}^{(0)}$ and $\widehat{\rho}_0 = 0$.

3.2 Effective degrees of freedom

After some manipulations of the expressions of $\widehat{\boldsymbol{\beta}}^{(k+1)}$ and $\widehat{\boldsymbol{\gamma}}^{(k+1)}$ in the **M-step**, we have that

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left[(\widehat{\mathbf{A}}^{(k)}\mathbf{X})^\top\widehat{\mathbf{W}}_x^{(k)}\widehat{\mathbf{A}}^{(k)}\mathbf{X}\right]^{-1}(\widehat{\mathbf{A}}^{(k)}\mathbf{X})^\top\widehat{\mathbf{W}}_x^{(k)}\widehat{\mathbf{y}}^{*(k)} \quad \text{and}\\
\widehat{\boldsymbol{\gamma}}^{(k+1)} &= \left[(\widehat{\mathbf{A}}^{(k)}\mathbf{N})^\top\widehat{\mathbf{W}}_f^{(k)}\widehat{\mathbf{A}}^{(k)}\mathbf{N} + \alpha\widehat{\sigma^2}^{(k)}\mathbf{K}\right]^{-1}(\widehat{\mathbf{A}}^{(k)}\mathbf{N})^\top\widehat{\mathbf{W}}_f^{(k)}\widehat{\mathbf{y}}^{*(k)},
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{W}_x &= \mathbf{I}_n - \mathbf{AN}[(\mathbf{AN})^\top\mathbf{AN} + \alpha\sigma^2\mathbf{K}]^{-1}(\mathbf{AN})^\top \quad \text{and}\\
\mathbf{W}_f &= \mathbf{I}_n - \mathbf{AX}[(\mathbf{AX})^\top\mathbf{AX}]^{-1}(\mathbf{AX})^\top.
\end{aligned}
$$

From Hastie and Tibshirani (1990) the effective degrees of freedom involved in modeling the nonparametric component will be estimated from the following relationship obtained at the convergence of $\widehat{\boldsymbol{\gamma}}^{(k+1)}$. Thus, we may express the estimate of the corrected linear predictor for the nonparametric component as

$$\widehat{\mathbf{A}\mathbf{N}}\widehat{\boldsymbol{\gamma}} = \widehat{\mathbf{H}}(\alpha)\widehat{\mathbf{y}}^{*},$$

where $\widehat{\mathbf{H}}(\alpha) = \widehat{\mathbf{A}\mathbf{N}}[(\widehat{\mathbf{A}\mathbf{N}})^{\top}\widehat{\mathbf{W}}_f\widehat{\mathbf{A}\mathbf{N}} + \alpha\widehat{\sigma^2}\mathbf{K}]^{-1}(\widehat{\mathbf{A}\mathbf{N}})^{\top}\widehat{\mathbf{W}}_f$ may be interpreted as a linear smoother. As pointed out by Hastie and Tibshirani (1990, p. 52), the sum of the eigenvalues of $\widehat{\mathbf{H}}(\alpha)$, for α fixed, namely

$$df(\alpha) = tr\{\widehat{\mathbf{H}}(\alpha)\}$$

may be defined as the effective degree of freedom due to the nonparametric fitting.

Then, one has a total of $p + 3 + df(\alpha)$ parameters to be estimated and we may use the Bayesian information criterion (BIC) to select an appropriate model, which consists in minimizing the function

$$BIC(\alpha) = -2\ell_p(\widehat{\boldsymbol{\theta}}, \alpha) + [p + 3 + df(\alpha)]\log(n),$$

where $\ell_p(\widehat{\boldsymbol{\theta}}, \alpha)$ denotes the penalized log-likelihood function available at $\widehat{\boldsymbol{\theta}}$ for a fixed α, given by $\ell_p(\widehat{\boldsymbol{\theta}}, \alpha) = \ell(\widehat{\boldsymbol{\theta}}) - \alpha\widehat{\boldsymbol{\gamma}}^{\top}\mathbf{K}\widehat{\boldsymbol{\gamma}}$. We use the “optim”routine in R (R Core Team 2019) to estimate α, with α between $(0, 10^3)$.

3.3 Residual analysis

We propose to use the quantile residuals (Dunn and Smyth 1996). From the model (4), the ith ordinary residual is given by

$$\begin{aligned} r_{q_i} &= y_i - \mathbf{x}_i^{\top}\widehat{\boldsymbol{\beta}} - \mathbf{n}_i^{\top}\widehat{\boldsymbol{\gamma}} - \widehat{\rho}\hat{\epsilon}_{i-1} \\ &= \begin{cases} y_1 - \mathbf{x}_1^{\top}\widehat{\boldsymbol{\beta}} - \mathbf{n}_1^{\top}\widehat{\boldsymbol{\gamma}}, & i = 1, \\ y_i - \mathbf{x}_i^{\top}\widehat{\boldsymbol{\beta}} - \mathbf{n}_i^{\top}\widehat{\boldsymbol{\gamma}} - \hat{\rho}\left[y_{i-1} - \mathbf{x}_{i-1}^{\top}\widehat{\boldsymbol{\beta}} - \mathbf{n}_{i-1}^{\top}\widehat{\boldsymbol{\gamma}}\right], & i = 2, \ldots, n. \end{cases} \end{aligned} \tag{12}$$

So, using the cdf of Y_i in (3), the conditional quantile residual is defined as

$$t_{q_i} = \Phi^{-1}(F_Y(r_{q_i}; -b\widehat{\delta}, \widehat{\sigma}^2, \widehat{\delta})), \quad i = 1, \ldots, n. \tag{13}$$

According to Dunn and Smyth (1996), the distribution of t_{q_i} converges to standard normal if the parameter $\boldsymbol{\theta}$ is consistently estimated. So, we will use conditional quantile residuals to the construction of simulated confidence bands for assessing departures from the error assumptions and the presence of outlying observations and construct graphs of the autocorrelation and partial autocorrelation functions for verifying the adequacy of the AR structure for the errors.

4 The local influence approach

The main idea of sensitivity studies is to assess changes in the parameter estimates, particularly inferential changes, under perturbations in the model or data. The most popular method is the case deletion, which consists in assessing changes in the parameter estimates after dropping individual observations (see, for instance, Cook and Weisberg 1982). This approach has been largely applied in regression models under independent observations, but has rarely appeared in time series models. The methodology proposed by Cook (1986), named local influence, which consists in studying the influence of small perturbations in the model or the data on the parameter estimates, seems more suitable for such models. However, the local influence approach was originally developed for maximum likelihood estimation through the likelihood displacement. Zhu and Lee (2001) proposed an extension of the methodology for the EM estimation covering a general class of statistical models with missing data. The approach is based on the Q-displacement instead of the likelihood displacement. We propose a natural extension of this methodology to semiparametric models.

Let $\mathbf{w} = (w_1, \ldots, w_n)^\top$ a perturbation vector varying in an open region $\boldsymbol{\Omega} \in \mathbb{R}^n$ and $\ell_{c_p}(\boldsymbol{\theta}, \mathbf{w}|\mathbf{y}_c)$, $\boldsymbol{\theta} \in \mathbb{R}^p$, the complete-data penalized log-likelihood function of the perturbed model and let $\widehat{\boldsymbol{\theta}}(\mathbf{w})$ denote the maximum of the function $Q_p(\boldsymbol{\theta}, \mathbf{w}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_{c_p}(\boldsymbol{\theta}, \mathbf{w}|\mathbf{Y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}]$. It is supposed that there is $\mathbf{w}_0$ such that $\ell_{c_p}(\boldsymbol{\theta}, \mathbf{w}_0|\mathbf{Y}_c) = \ell_{c_p}(\boldsymbol{\theta}|\mathbf{Y}_c)$ for all $\boldsymbol{\theta}$. So, the influence graph is defined as $\boldsymbol{\alpha}(\mathbf{w}) = (\mathbf{w}^\top, f_Q(\mathbf{w}))^\top$, where $f_Q(\mathbf{w})$ is the Q-displacement function defined as

$$f_Q(\mathbf{w}) = 2\left[Q_p\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q_p\left(\widehat{\boldsymbol{\theta}}(\mathbf{w})|\widehat{\boldsymbol{\theta}}\right)\right].$$

According to Zhu and Lee (2001), the normal curvature $C_{f_Q,\mathbf{d}}$, of $\boldsymbol{\alpha}(\mathbf{w})$ available at $\mathbf{w}_0$ in the direction of some unit vector $\mathbf{d}$ has information about the local behavior of the Q-displacement function. Zhu and Lee (2001) show that

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\mathbf{w}_o}\mathbf{d} \quad \text{and} \quad -\ddot{Q}_{\mathbf{w}_0} = \Delta_{\mathbf{w}_o}^\top \left\{-\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})\right\}^{-1}\Delta_{\mathbf{w}_0},$$

where $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = \left.\dfrac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\Delta_{\mathbf{w}} = \left.\dfrac{\partial^2 Q_p(\boldsymbol{\theta}, \mathbf{w}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\mathbf{w}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\mathbf{w})}$.

To construct a measure of influence, we calculate the spectral decomposition of $-\ddot{Q}_{\mathbf{w}_0}$

$$-2\ddot{Q}_{\mathbf{w}_o} = \sum_{k=1}^{n} \xi_k \mathbf{e}_k \mathbf{e}_k',$$

where $(\xi_1, \mathbf{e}_1), \ldots, (\xi_n, \mathbf{e}_n)$ are the eigenvalue-eigenvector pairs of the matrix $-2\ddot{Q}_{\mathbf{w}_o}$ with $\xi_1 \geq \cdots \geq \xi_q, \xi_{q+1} = \cdots = \xi_n = 0$ and $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are elements of the associated orthonormal basis. Following Zhu and Lee (2001) and Lu and Song (2006), we consider an aggregated contribution vector of all eigenvectors corresponding to nonzero eigenvalues, given by

$$M(0)_l = \sum_{k=1}^{q} \widetilde{\xi}_k \mathbf{e}_{kl}^2,$$

where $\widetilde{\xi}_k = \xi_k / \sqrt{\sum_{j=1}^{q} \xi_j^2}$ and $\mathbf{e}_k^2 = (e_{k1}^2, \ldots, e_{kn}^2)$.

Hence, we inspect the graphic of $\{M(0)_l, l = 1, \ldots, n\}$ to the assessment of influential cases. Following Lee and Xu (2004), we use the value $1/n + c^* S$ as a benchmark to regard the lth case as influential, where c^* is a selected constant (depending on the real application) and S is the standard deviation of the vector $\{M(0)_l, l = 1, \ldots, n\}$.

4.1 The Hessian matrix, $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})$

To obtain the diagnostic measures of the PLM-SNAR(1), based on the approach of Zhu and Lee (2001), it is necessary to compute $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \sigma^2, \delta, \rho)^\top$. It follows from (8) to (9) that $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})$ has elements given by

$$
\begin{aligned}
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= -\frac{1}{\sigma^2} (\mathbf{AX})^\top (\mathbf{AX}), \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^\top} &= -\frac{1}{\sigma^2} (\mathbf{AN})^\top (\mathbf{AX}), \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \boldsymbol{\beta}} &= -\frac{1}{\sigma^4} (\mathbf{AX})^\top \left[\mathbf{y} - \boldsymbol{\xi} - \delta(\widehat{\mathbf{z}} - b\mathbf{1}_n) \right], \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \delta \partial \boldsymbol{\beta}} &= -\frac{1}{\sigma^2} (\mathbf{AX})^\top (\widehat{\mathbf{z}} - b\mathbf{1}_n), \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \rho \partial \boldsymbol{\beta}} &= -\frac{1}{\sigma^2} \sum_{i=2}^{n} \left[(y_i - \xi_i - \delta(\widehat{z}_i - b)) \mathbf{x}_{i-1} + r_{i-1} (\mathbf{x}_i - \rho \mathbf{x}_{i-1}) \right], \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} &= -\frac{1}{\sigma^2} (\mathbf{AN})^\top (\mathbf{AN}) - \alpha \mathbf{K}, \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \boldsymbol{\gamma}} &= -\frac{1}{\sigma^4} (\mathbf{AN})^\top \left[\mathbf{y} - \boldsymbol{\xi} - \delta(\widehat{\mathbf{z}} - b\mathbf{1}_n) \right], \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \delta \partial \boldsymbol{\gamma}} &= -\frac{1}{\sigma^2} (\mathbf{AN})^\top (\widehat{\mathbf{z}} - b\mathbf{1}_n), \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \rho \partial \boldsymbol{\gamma}} &= -\frac{1}{\sigma^2} \sum_{i=2}^{n} \left[(y_i - \xi_i - \delta(\widehat{z}_i - b)) \mathbf{n}_{i-1} + r_{i-1} (\mathbf{n}_i - \rho \mathbf{n}_{i-1}) \right], \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \sigma^4} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^{n} \left[(y_i - \xi_i)^2 - 2\delta (y_i - \xi_i)(\widehat{z}_i - b) + \delta^2 (b^2 - 2b\widehat{z}_i + \widehat{z^2}_i) \right], \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \delta \partial \sigma^2} &= \frac{1}{\sigma^4} \sum_{i=1}^{n} \left[-(y_i - \xi_i)(\widehat{z}_i - b) + \delta (b^2 - 2b\widehat{z}_i + \widehat{z^2}_i) \right], \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \rho \partial \sigma^2} &= -\frac{1}{\sigma^4} \sum_{i=2}^{n} \left[r_i - \rho r_{i-1} - \delta(\widehat{z}_i - b) \right] r_{i-1},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \delta^2} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}(b^2 - 2b\widehat{z}_i + \widehat{z^2}_i),\\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \rho \partial \delta} &= -\frac{1}{\sigma^2}\sum_{i=2}^{n}(\widehat{z}_i - b)r_{i-1} \text{ and}\\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \rho^2} &= -\frac{1}{\sigma^2}\sum_{i=2}^{n} r_{i-1}^2,
\end{aligned}
$$

where $\mathbf{1}_n$ is a one's vector and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^\top$.

4.2 Perturbation schemes

In this section, we consider the three usual perturbation schemes in local influence for the PLM-SNAR(1) proposed in this work.

4.2.1 Case-weight perturbation

Under this scheme, we evaluate if the contributions of the observations with different weights affect the ML estimate of $\boldsymbol{\theta}$. The perturbed Q-function is written as

$$
Q_p(\boldsymbol{\theta}, \mathbf{w}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} w_i Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) - \frac{\alpha}{2}\boldsymbol{\gamma}^\top \mathbf{K}\boldsymbol{\gamma}.
$$

In this case, $\mathbf{w}_0 = (1, \ldots, 1)^\top = \mathbf{1}_n$ and $\dfrac{\partial Q_p(\boldsymbol{\theta}, \mathbf{w}|\widehat{\boldsymbol{\theta}})}{\partial \mathbf{w}_i} = Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ and $\boldsymbol{\Delta}_{\mathbf{w}_0}$ has elements $\dfrac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}$, $i = 1, \ldots, n$, given by

$$
\begin{aligned}
\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\left[y_i - \xi_i - \delta(\widehat{z}_i - b)\right](\mathbf{x}_i - \rho\mathbf{x}_{i-1}),\\
\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma}} &= \frac{1}{\sigma^2}\left[y_i - \xi_i - \delta(\widehat{z}_i - b)\right](\mathbf{n}_i - \rho\mathbf{n}_{i-1}),\\
\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\left[(y_i - \xi_i)^2 - 2\delta(y_i - \xi_i)(\widehat{z}_i - b) + \delta^2(b^2 - 2b\widehat{z}_i + \widehat{z^2}_i)\right],\\
\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \delta} &= \frac{1}{\sigma^2}\left[(y_i - \xi_i)(\widehat{z}_i - b) - \delta(b^2 - 2b\widehat{z}_i + \widehat{z^2}_i)\right] \text{ and}\\
\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \rho} &= \frac{1}{\sigma^2}\left[r_i - \rho r_{i-1} - \delta(\widehat{z}_i - b)\right] r_{i-1},
\end{aligned}
$$

with $\mathbf{x}_0 = \mathbf{n}_0 = \mathbf{0}$ and $r_0 = 0$.

4.2.2 Explanatory variable perturbation

Considering a specific continuous explanatory variable, we use an additive perturbation given by

$$\mathbf{x}_{r_{\mathbf{w}}} = \mathbf{x}_r + S_r \mathbf{w}, \quad r \in \{1, \ldots, p\},$$

where S_r is the standard deviation of the explanatory variable $\mathbf{x}_r$. So, $\mathbf{w}_0 = \mathbf{0} : n \times 1$ and the perturbed Q-function is given by

$$Q_p(\boldsymbol{\theta}, \mathbf{w}|\widehat{\boldsymbol{\theta}}) \propto -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(y_i - \xi_{iw})^2 - 2\delta(y_i - \xi_{iw})(\widehat{z_i} - b)\right],$$

where $\xi_{iw} = \begin{cases} \mathbf{x}_{1w}^\top \boldsymbol{\beta} + \mathbf{n}_1^\top \boldsymbol{\gamma}, & i = 1, \\ (\mathbf{x}_{iw} - \rho\mathbf{x}_{i-1,w})^\top \boldsymbol{\beta} + (\mathbf{n}_i - \rho\mathbf{n}_{i-1})^\top \boldsymbol{\gamma} + \rho y_{i-1}, & i = 2, \ldots, n. \end{cases}$

It follows that the matrix $\boldsymbol{\Delta}_{\mathbf{w}_0} = \left.\dfrac{\partial^2 Q_p(\boldsymbol{\theta}, \mathbf{w}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\mathbf{w}^\top}\right|_{\mathbf{w}=\mathbf{w}_0} = \left(\Delta_{\boldsymbol{\beta}}^\top, \Delta_{\sigma^2}^\top, \Delta_{\lambda}^\top, \Delta_{\boldsymbol{\gamma}}^\top\right)^\top,$ where

$$\Delta_{\boldsymbol{\beta}_i} = \frac{S_r}{\sigma^2}\begin{cases} \left[y_i - \xi_i - \delta(\widehat{z_i} - b) - \rho(y_{i+1} - \xi_{i+1} - \delta(\widehat{z}_{i+1} - b))\right]\mathbf{1}_{r0} \\ -\beta_r\left[\mathbf{x}_i - \rho\mathbf{x}_{i-1} - (\mathbf{x}_{i+1} - \rho\mathbf{x}_i)\right], \quad i = 1, \ldots, n-1, \\ (y_n - \xi_n - \delta(\widehat{z}_n - b))\mathbf{1}_{r0} - \beta_r(\mathbf{x}_n - \rho\mathbf{x}_{n-1}), \quad i = n, \end{cases}$$

$$\Delta_{\boldsymbol{\gamma}_i} = -\frac{S_r\beta_r}{\sigma^2}\begin{cases} \mathbf{n}_i - \rho\mathbf{n}_{i-1} - (\mathbf{n}_{i+1} - \rho\mathbf{n}_i), & i = 1, \ldots, n-1, \\ \mathbf{n}_n - \rho\mathbf{n}_{n-1}, & i = n, \end{cases}$$

$$\Delta_{\sigma^2{}_i} = \frac{S_r\beta_r}{\sigma^4}\begin{cases} y_i - \xi_i - \delta(\widehat{z_i} - b) - \rho(y_{i+1} - \xi_{i+1} - \delta(\widehat{z}_{i+1} - b)), & i = 1, \ldots, n-1. \\ y_n - \xi_n - \delta(\widehat{z}_n - b), & i = n, \end{cases}$$

$$\Delta_{\delta_i} = -\frac{S_r\beta_r}{\sigma^2}\begin{cases} \widehat{z_i} - b - \rho(\widehat{z}_{i+1} - b), & i = 1, \ldots, n-1, \\ \widehat{z}_n - b, & i = n, \end{cases}$$

$$\Delta_{\rho_i} = \frac{S_r\beta_r}{\sigma^2}\begin{cases} \mu_{i-1} - y_{i-1} - \rho(\mu_i - y_i) - y_{i+1} + \xi_{i+1} + \delta(\widehat{z}_{i+1} - b), & i = 1, \ldots, n-1, \\ \mu_{i-1} - y_{n-1} \quad i = n, \end{cases}$$

where β_r is the rth element of $\boldsymbol{\beta}$ and $\mathbf{1}_{r0}$ is a $p \times 1$ vector with one in the tth position and zeros elsewhere.

4.2.3 Response variable perturbation

We consider an additive perturbation given by

$$y_{iw} = y_i + S_y w_i, \quad i = 1, \ldots, n,$$

where S_y is the standard deviation of $\mathbf{y}$. In this case, $\mathbf{w}_0 = \mathbf{0} : n \times 1$ and

$$Q_p(\boldsymbol{\theta}, w_i|\widehat{\boldsymbol{\theta}}) \propto -\frac{1}{2\sigma^2}\left[(y_{iw} - \xi_{iw})^2 - 2\delta(y_{iw} - \xi_{iw})(\widehat{z_i} - b)\right]$$

$$-\frac{1}{2\sigma^2}\left[(y_{i+1,w}-\xi_{i+1,w})^2-2\delta(y_{i+1,w}-\xi_{i+1,w})(\widehat{z}_{i+1}-b)\right],$$

where $\xi_{iw} = \begin{cases} \mathbf{x}_1^\top\boldsymbol{\beta}+\mathbf{n}_1^\top\boldsymbol{\gamma}, & i=1, \\ (\mathbf{x}_i-\rho\mathbf{x}_{i-1})^\top\boldsymbol{\beta}+(\mathbf{n}_i-\rho\mathbf{n}_{i-1})^\top\boldsymbol{\gamma}+\rho y_{i-1,w}, & i=2,\ldots,n. \end{cases}$

It follows that the matrix $\boldsymbol{\Delta}_{\mathbf{w}_0} = \left.\dfrac{\partial^2 Q_p(\boldsymbol{\theta},\mathbf{w}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\mathbf{w}^\top}\right|_{\mathbf{w}=\mathbf{w}_0} = \left(\Delta_{\boldsymbol{\beta}}^\top, \Delta_{\sigma^2}^\top, \Delta_\lambda^\top, \Delta_{\boldsymbol{\gamma}}^\top\right)^\top,$ where

$$\Delta_{\boldsymbol{\beta}_i} = \frac{S_y}{\sigma^2}\begin{cases} (1+\rho^2)\mathbf{x}_1-\rho\mathbf{x}_2, & i=1, \\ \rho(\mathbf{x}_{i+1}-\rho\mathbf{x}_i)\left[-1+\delta(\widehat{z}_{i+1}-b)\right]+\mathbf{x}_i-\rho\mathbf{x}_{i-1}, & i=2,\ldots,n-1, \\ \mathbf{x}_n-\rho\mathbf{x}_{n-1}, & i=n, \end{cases}$$

$$\Delta_{\boldsymbol{\gamma}_i} = \frac{S_y}{\sigma^2}\begin{cases} (1+\rho^2)\mathbf{n}_1-\rho\mathbf{n}_2, & i=1, \\ \rho(\mathbf{n}_{i+1}-\rho\mathbf{n}_i)\left[-1+\delta(\widehat{z}_{i+1}-b)\right]+\mathbf{n}_i-\rho\mathbf{n}_{i-1}, & i=2,\ldots,n-1, \\ \mathbf{n}_n-\rho\mathbf{n}_{n-1}, & i=n, \end{cases}$$

$$\Delta_{\sigma^2_i} = \frac{S_y}{\sigma^4}\begin{cases} y_1-\mu_1-\delta(\widehat{z}_1-b)-\rho(y_2-\xi_2)+\rho\delta(\widehat{z}_2-b), & i=1, \\ \rho(y_{i+1}-\xi_{i+1})\left[-1+\delta(\widehat{z}_{i+1}-b)\right]+y_i-\xi_i-\delta(\widehat{z}_i-b), & i=2,\ldots,n-1, \\ y_n-\xi_n-\delta(\widehat{z}_n-b), & i=n, \end{cases}$$

$$\Delta_{\delta_i} = \frac{S_y}{\sigma^2}\begin{cases} \widehat{z}_1-b-\rho(\widehat{z}_2-b), & i=1, \\ \widehat{z}_i-b-\rho(y_{i+1}-\xi_{i+1})(\widehat{z}_{i+1}-b), & i=2,\ldots,n-1, \\ \widehat{z}_n-b, & i=n, \end{cases}$$

$$\Delta_{\rho_i} = \frac{S_y}{\sigma^2}\begin{cases} y_2-\xi_2-\delta(\widehat{z}_2-b)-\rho(y_1-\mu_1), & i=1, \\ \left[\rho(y_i-\mu_i)-y_{i+1}+\xi_{i+1}\right]\left[-1+\delta(\widehat{z}_{i+1}-b)\right]+y_{i-1}-\mu_{i-1}, & i=2,\ldots,n-1, \\ y_n-\mu_n, & i=n. \end{cases}$$

5 Simulation study

In order to study the large sample behavior of the natural cubic spline estimate obtained from the iterative described in Sect. 3.1, we perform a small simulation study in which the true nonparametric function assumes the forms: $f_1(t)=\cos(t), t\in(-\pi/2, 4\pi)$ (model 1, coseno) or $f_2(t)=cos(4\pi t)exp(-t^2/2), t\in(0.6, 1.6)$ (model 2, doppler effect).

In addition, the PLM-SNAR(1) is expressed as follows

$$y_i=\beta x_i+f(t_i)+\epsilon_i, \quad i=1,\ldots,n,$$

with $\beta=5$, $\sigma^2=0.005$, $\delta=0.212$, $\rho=0.8$ where x_i are generated from a $U(0,1)$. In the generation of $f_1(t)$ and $f_2(t)$, we divide the interval of t in n points equally spaced. The simulation study is performed for different sample sizes (n equals 100, 500 and 1000) and $M=1000$ replicates for each situation. For each random sample, we calculate the parameter estimates using the EM-algorithm in Sect. 3 and the approximate standard errors using the observed information matrix ("Appendix A"). With these estimates, we calculated for each parameter θ_k (β, σ^2, δ and ρ) the mean of the estimates (*Mean*) $\bar{\widehat{\theta}}_k=\sum_{j=1}^{M}\widehat{\theta}_{k_j}/M$, the standard deviation of the estimates (*SD*) given by $\sqrt{\sum_{j=1}^{M}(\widehat{\theta}_{k_j}-\bar{\widehat{\theta}}_k)^2/(M-1)}$ and the empirical standard errors (SE_{emp})

Table 1 Mean and standard deviations (SD) for EM estimates and empirical standard error estimates (SE_{emp}) based on 1000 samples from the model 1 (coseno)

Parameter	True Value	n = 100 Mean	SD	SE_{emp}	n = 500 Mean	SD	SE_{emp}	n = 1000 Mean	SD	SE_{emp}
β	5.000	5.000	0.045	0.043	5.000	0.017	0.016	5.000	0.011	0.011
σ^2	0.005	0.010	0.006	0.003	0.005	0.004	0.001	0.005	0.002	0.001
δ	0.212	0.073	0.092	0.827	0.199	0.050	0.038	0.208	0.025	0.013
ρ	0.800	0.394	0.129	0.103	0.751	0.031	0.028	0.776	0.019	0.018

Table 2 Mean and standard deviations (SD) for EM estimates and empirical standard error estimates (SE_{emp}) based on 1000 samples from the model 2 (doppler effect)

Parameter	True Value	n = 100 Mean	SD	SE_{emp}	n = 500 Mean	SD	SE_{emp}	n = 1000 Mean	SD	SE_{emp}
β	5.000	4.998	0.042	0.040	5.000	0.016	0.016	5.000	0.011	0.011
σ^2	0.005	0.005	0.003	0.003	0.005	0.004	0.001	0.005	0.001	0.001
δ	0.212	0.074	0.093	0.556	0.200	0.050	0.036	0.209	0.022	0.012
ρ	0.800	0.411	0.111	0.097	0.751	0.030	0.027	0.776	0.019	0.018

calculated as $\sqrt{\sum_{j=1}^{M} I_j^{kk}/M}$, where I_j^{kk} denotes the k-th diagonal element of an observed information matrix inverse. When sample size increases, it is expected that the bias $(\widehat{\theta_k} - \theta_k)$ decreases and the values SD and SE_{emp} get smaller and closer.

Tables 1 and 2 present the results of the simulation study for the parametric components for models 1 and 2. As expected, the bias of the MLE decreases as the sample size increases. In addition, the standard deviation of the estimates as well as the empirical standard errors are closer to each other and decrease when the sample size increases, indicating that the observed information matrix calculated seems to be correct (Fig. 1).

6 Application

The data consist of daily measurements of Ozone concentration (maximum 1 h average) and Temperature (the Inversion base temperature, degrees Fahrenheit) in Upland, California, U.S.A., in 1976 (Breiman and Friedman 1985). The main interest is to explain the daily Ozone concentration (ppm). Figure 2a and b present the histogram and boxplot for the Ozone concentration, which seems to follow some suitable behaviour, that could be fitted, for instance, by asymmetric distributions such as skew-normal. Figure 2c presents the scatter-plot of Ozone and Temperature which seems a linear relation. Besides, Fig. 2d presents the time series of the Ozone concentration, where we may notice a nonlinear behaviour, which can be modeled by a partially linear model.

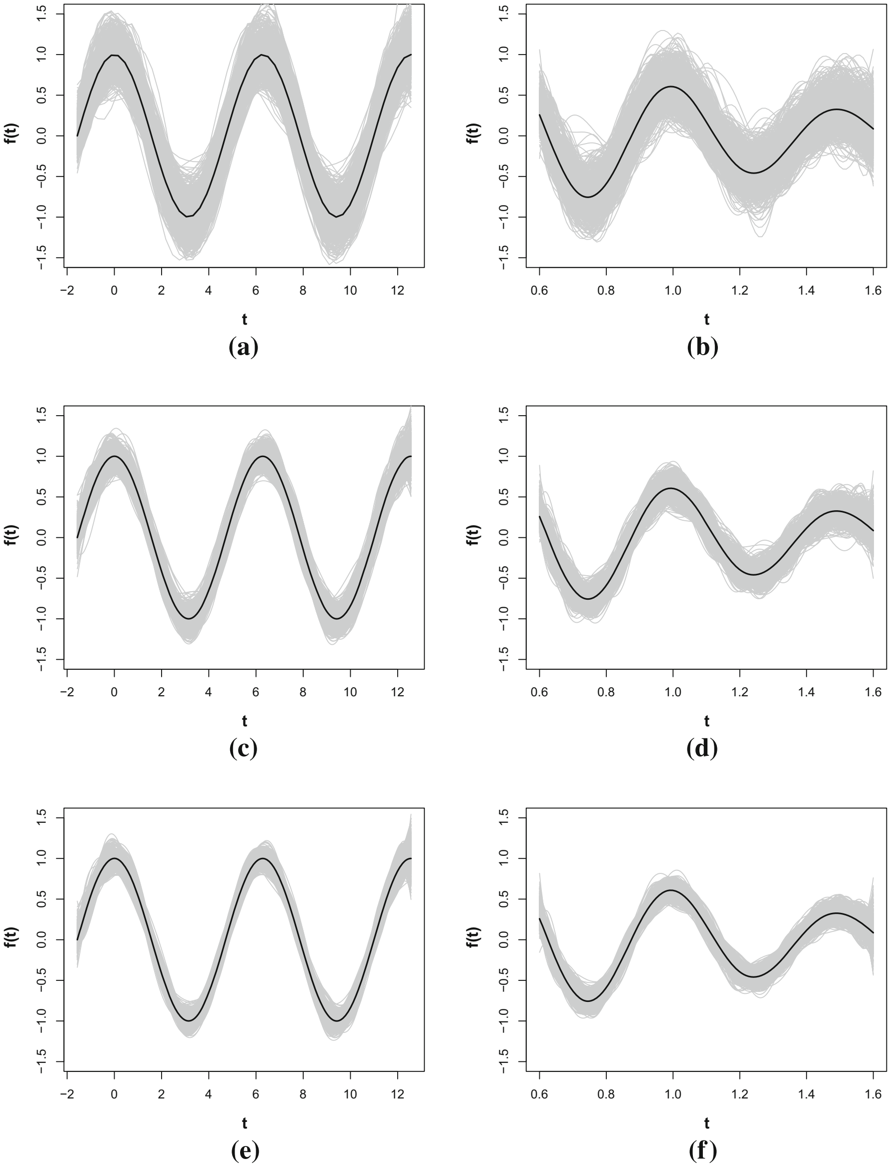


Fig. 1 Graphs of the nonparametric components with 1000 replications. Adjusted curves (gray lines) and true curves (black lines): model 1 (first column) and model 2 (second column). **a** and **b** for $n = 100$; **c** and **d** for $n = 500$, **e** and **f** for $n = 1000$

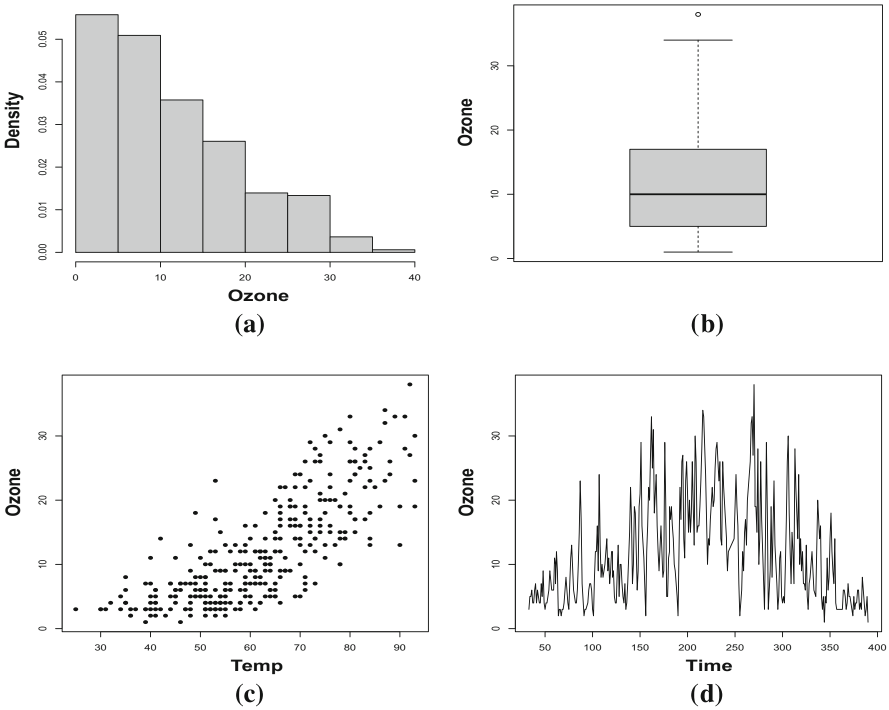


Fig. 2 Graphics for pollen concentration: **a** histogram of the original scale; histogram **b**, boxplot **c** and **d** time series for the Ozone concentration

So, we propose the following partially linear model:

$$y_i = \beta_0 + \beta_1 x_i + f(t_i) + \epsilon_i, \quad i = 1, \ldots, 330, \tag{14}$$

where y_i and x_i denote, respectively, the Ozone concentration (in ppm) and the temperature at the i-th day t_i, $f(t_i)$ is an arbitrary smooth function and $\epsilon_i \sim N(0, \sigma^2)$ are independent errors.

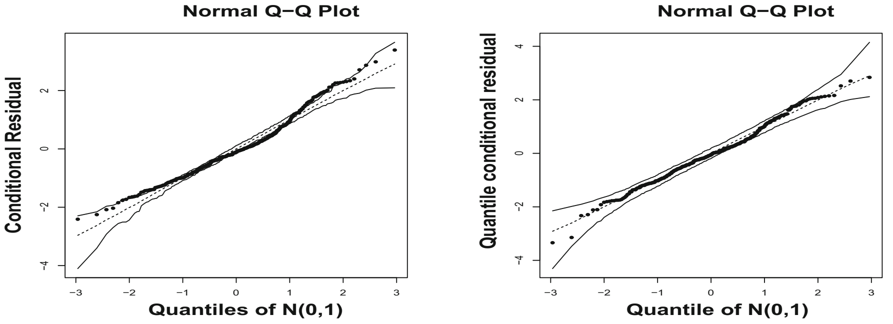
Under model (14), we noticed departures of the quantile conditional residual from a white noise process. The ACF of the partial residuals presents exponential decay while the PACF was significantly different from zero only in the first lag, indicating a behaviour typical of AR(1) process. Thus, in order to take into account the periodic tendency as well as the temporal correlation among the observations, we will consider the same model given in (14), but by assuming now AR(1) errors. The residuals also presented signs of skewness, which suggests the use of a skewed distribution to model the errors.

Based on the earlier analysis, we will propose a partially linear model with AR(1)-Skew-normal errors to the model (14). For the purpose of comparison, we will also fit the model under AR(1)-normal error, named PLM-NAR(1).

Table 3 presents the parameter estimates, approximate standard errors using the observed information matrix ("Appendix A"), degrees of freedom and the BIC values

Table 3 Maximum penalized likelihood estimates (S.E. denotes approximate standard errors) under AR(1)-Normal and AR(1)-Skew-normal partially linear models fitted to the Ozone concentration

	AR(1)-Normal		AR(1)-Skew-normal	
Effect	Estimate	SE	Estimate	SE
Intercept	− 11.838	0.137	− 7.324	0.099
Temp.	0.382	0.002	0.311	0.002
σ^2	20.278	2.492	3.648	0.591
δ	–		6.802	0.103
ρ	0.330	0.003	0.333	0.003
α	1.000		0.999	
$df(\alpha)$	4.267		5.269	
$\ell_p(\widehat{\boldsymbol{\theta}}, \alpha)$	− 966.184		− 949.665	
BIC	1980.310		1958.880	



(a) **(b)**

Fig. 3 Quantile-quantile plots with a 99% simulated confidence band for the conditional quantile residuals from the fit of the partially linear model with AR(1)-normal (**a**) and AR(1)-Skew-normal (**b**) errors to the data set on Ozone concentration

for the fitted models. Comparing such estimates we may notice that all parametric effects are significant with a good agreement among the fitted models. Based on the BIC we may observe a little better goodness-of-fit of the PLM-SNAR(1) model.

Figure 3a and b present the quantile-quantile plots with a simulated confidence band of 99% (Atkinson 1981) based on the conditional quantile residuals t_{q_i}, for $i = 1, \ldots, n$, for Normal and Skew-normal models, respectively. We see that the PLM-SNAR(1) does not present observations out of the confidence band (Fig. 3b), whereas the PLM-NAR(1) presents various observations outside (Fig. 3a).

So, based on the comparison of the goodness-of-fit and graphics of the two models, the PLM-SNAR(1) seems to have a better performance to explain the Ozone concentration. Next, we present other complementary analysis as well as the local influence analyses for the PLM-SNAR(1).

Figure 4a presents the fitted nonparametric function, the partial residuals $(y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}})$ and the 99% pointwise confidence band for the nonparametric function, whereas Fig. 4b presents the conditional quantile residuals, which indicate no observed tendency. The autocorrelation and partial autocorrelation functions for the conditional

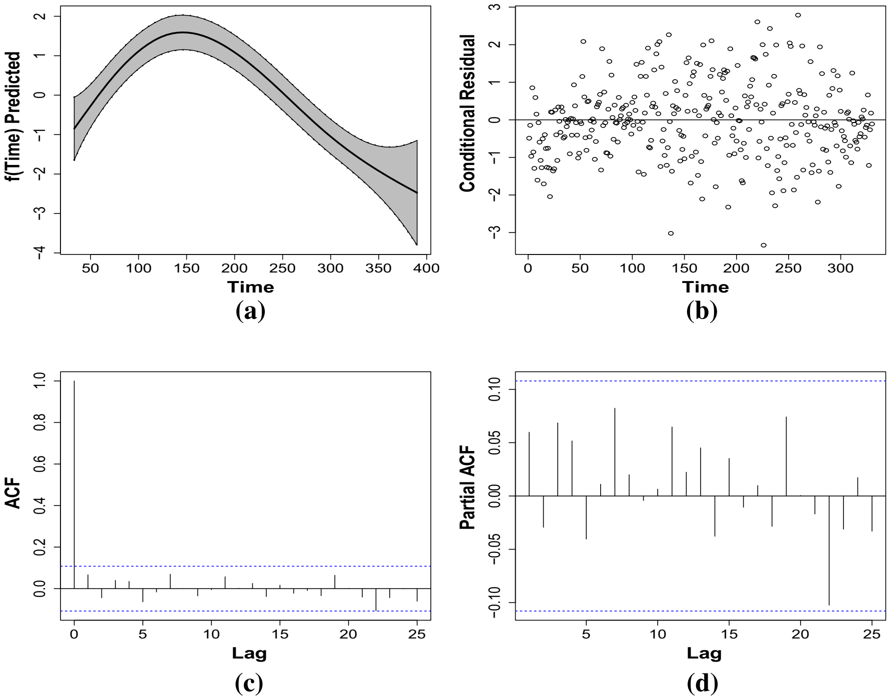
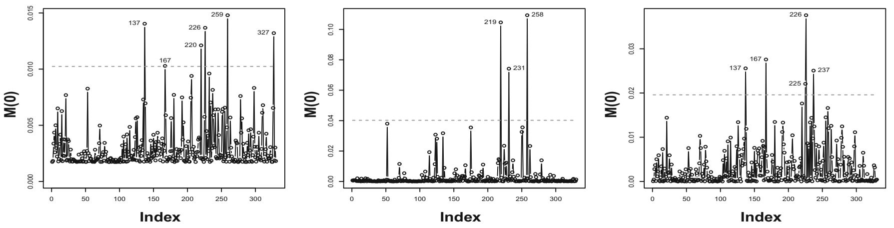


Fig. 4 Fit of the partially linear model with AR(1)-Skew-normal errors to the data set on Ozone concentration: Partial residuals and approximate 99% pointwise confidence band for the nonparametric function (**a**); Graphics of the conditional quantile residual: by time (**b**), Autocorrelation function (**c**) and (**d**) Partial autocorrelation function

quantile residual t_{q_i} displayed in the Fig. 4c and d, respectively, indicate that the AR(1) structure assumed for the errors is sufficient to contemplate the temporal correlation among the observations.

Next, we conduct a local influence study based on $M(0)_l$ from the conformal curvature B_i, obtained by considering the case-weight, response and the explanatory variable perturbation schemes. Figure 5a–c present, respectively, the index plots of $M(0)_l$ for the PLM-SNAR(1) model for case-weight, response and explanatory (Temperature) perturbations. For case-weight perturbation, observations #137, #167, #220, #226, #259 and #327 appear as possible influential that are high residuals in absolute value, being #137 and #259 the smallest and the highest residuals, respectively. For the response perturbation, the observation #219 has high Temperature, while #231 and #258 presents median values for Ozone and Temperature, and moderate residuals. The influential observations in the Temperature perturbation, #137, #167, #225, #226, #237 and #327, present high absolute residuals.



(a) **(b)** **(c)**

Fig. 5 Diagnostic graphs from the PLM-SNAR(1) fitted to the data set on Ozone concentration with $c^* = 3$: **a** case-weight perturbation; **b** response perturbation; and **c** explanatory (temperature) perturbation

7 Final considerations

In this paper, we discuss parameter estimation and some statistical diagnostics for partially linear models with first-order autoregressive error under a skew-normal distribution. We developed an EM algorithm for estimating the parametric and nonparametric components of the model with analytic expressions for M and E-steps. An appropriate smoother is also derived for estimating the effective degree of freedom concerned with the nonparametric fitting. Local influence approaches using Zhu and Lee's approach (Zhu and Lee 2001) were developed for the proposed model under case-weight, explanatory and response variables perturbations. We also derived closed-forms expressions for the observed information matrix, for the Hessian matrix $\ddot{\mathbf{Q}}$ and for the $\boldsymbol{\Delta}$ matrix. A simulation study considering two different nonlinear functions for the nonparametric component suggests that, for a large sample, the EM estimators seem unbiased, their variances may be estimated by using the inverse of the observed information matrix and one indicates the consistency of the nonparametric estimator. The model is applied to the daily ozone concentration in order to illustrate the application of the proposed methodology, showing the usefulness of PLM-SNAR(1) to fit data sets with nonparametric components in which the responses are asymmetric with dependence on time. The R codes (R Core Team 2019) developed in this work may be available upon request. A future work is to propose a PLM-AR(1) using an appropriate skew heavier-tailed family of distributions, such as SSMN (Ferreira et al. 2011).

Acknowledgements We thank two anonymous referees for comments that improved this manuscript. This work was partially supported by CNPq-Brazil and FAPEMIG-Brazil.

Appendix A: Approximate standard errors

The variance–covariance matrix of the $\widehat{\boldsymbol{\theta}}$, corresponding to the inverse of the observed information matrix, is obtained by treating the penalized likelihood as a usual likelihood Segal et al. (1994). Given the observed log-likelihood function $\ell(\boldsymbol{\theta})$ in (5), the correspondence penalized log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \sigma^2, \delta, \rho)^\top$ is of the form

$$\ell_p(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \frac{\alpha}{2}\boldsymbol{\gamma}^\top \mathbf{K}\boldsymbol{\gamma}. \tag{15}$$

The observed information matrix for $\boldsymbol{\theta}$ may be written as

$$\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} = -\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$$

with elements

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = \sum_{i=1}^{n}\left(-\frac{\sigma^2}{\delta^2} + W'_\Phi(B_i)\right)\frac{\partial B_i}{\partial\boldsymbol{\beta}}\frac{\partial B_i}{\partial\boldsymbol{\beta}^\top},$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\beta}^\top} = \sum_{i=1}^{n}\left(-\frac{\sigma^2}{\delta^2} + W'_\Phi(B_i)\right)\frac{\partial B_i}{\partial\boldsymbol{\gamma}}\frac{\partial B_i}{\partial\boldsymbol{\beta}^\top},$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top} = \sum_{i=1}^{n}\left(-\frac{\sigma^2}{\delta^2} + W'_\Phi(B_i)\right)\frac{\partial B_i}{\partial\boldsymbol{\gamma}}\frac{\partial B_i}{\partial\boldsymbol{\gamma}^\top} - \alpha\mathbf{K},$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\rho^2} = \sum_{i=1}^{n}\left(-\frac{\sigma^2}{\delta^2} + W'_\Phi(B_i)\right)\left(\frac{\partial B_i}{\partial\rho}\right)^2,$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\rho\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\left[\left(-\frac{\sigma^2}{\delta^2} + W'_\Phi(B_i)\right)\frac{\partial B_i}{\partial\rho}\frac{\partial B_i}{\partial\boldsymbol{\beta}} + \left(-\frac{\sigma^2}{\delta^2}B_i + W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\rho\partial\boldsymbol{\beta}}\right],$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\rho\partial\boldsymbol{\gamma}} = \sum_{i=1}^{n}\left[\left(-\frac{\sigma^2}{\delta^2} + W'_\Phi(B_i)\right)\frac{\partial B_i}{\partial\rho}\frac{\partial B_i}{\partial\boldsymbol{\gamma}} + \left(-\frac{\sigma^2}{\delta^2}B_i + W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\rho\partial\boldsymbol{\gamma}}\right],$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\sigma^2\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\left[\left(-\frac{1}{\delta^2}B_i - \frac{\sigma^2}{\delta^2}\frac{\partial B_i}{\partial\sigma^2} + W'_\Phi(B_i)\frac{\partial B_i}{\partial\sigma^2}\right)\frac{\partial B_i}{\partial\boldsymbol{\beta}} + \left(-\frac{\sigma^2}{\delta^2}B_i + W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\sigma^2\partial\boldsymbol{\beta}}\right],$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\sigma^2\partial\boldsymbol{\gamma}} = \sum_{i=1}^{n}\left[\left(-\frac{1}{\delta^2}B_i - \frac{\sigma^2}{\delta^2}\frac{\partial B_i}{\partial\sigma^2} + W'_\Phi(B_i)\frac{\partial B_i}{\partial\sigma^2}\right)\frac{\partial B_i}{\partial\boldsymbol{\gamma}} + \left(-\frac{\sigma^2}{\delta^2}B_i + W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\sigma^2\partial\boldsymbol{\gamma}}\right],$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\sigma^2\partial\rho} = \sum_{i=1}^{n}\left[\left(-\frac{1}{\delta^2}B_i - \frac{\sigma^2}{\delta^2}\frac{\partial B_i}{\partial\sigma^2} + W'_\Phi(B_i)\frac{\partial B_i}{\partial\sigma^2}\right)\frac{\partial B_i}{\partial\rho} + \left(-\frac{\sigma^2}{\delta^2}B_i + W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\sigma^2\partial\rho}\right],$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\delta\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\left[\left(\frac{2\sigma^2}{\delta^3}B_i - \frac{\sigma^2}{\delta^2}\frac{\partial B_i}{\partial\delta} + W'_\Phi(B_i)\frac{\partial B_i}{\partial\delta}\right)\frac{\partial B_i}{\partial\boldsymbol{\beta}} + \left(-\frac{\sigma^2}{\delta^2}B_i + W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\delta\partial\boldsymbol{\beta}}\right],$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\delta\partial\boldsymbol{\gamma}} = \sum_{i=1}^{n}\left[\left(\frac{2\sigma^2}{\delta^3}B_i - \frac{\sigma^2}{\delta^2}\frac{\partial B_i}{\partial\delta} + W'_\Phi(B_i)\frac{\partial B_i}{\partial\delta}\right)\frac{\partial B_i}{\partial\boldsymbol{\gamma}} + \left(-\frac{\sigma^2}{\delta^2}B_i + W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\delta\partial\mathbf{f}}\right],$$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\delta\partial\rho} = \sum_{i=1}^{n}\left[\left(\frac{2\sigma^2}{\delta^3}B_i - \frac{\sigma^2}{\delta^2}\frac{\partial B_i}{\partial\delta} + W'_\Phi(B_i)\frac{\partial B_i}{\partial\delta}\right)\frac{\partial B_i}{\partial\rho} + \left(-\frac{\sigma^2}{\delta^2}B_i + W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\delta\partial\rho}\right],$$

$$\begin{aligned}\frac{\partial^2 \ell_p(\boldsymbol{\theta})}{\partial\sigma^4} = {} & \frac{n}{2(\sigma^2+\delta^2)^2} - \frac{1}{\delta^2}\sum_{i=1}^{n}B_i\frac{\partial B_i}{\partial\sigma^2} \\ & + \sum_{i=1}^{n}\left[\left(-\frac{1}{\delta^2}B_i - \frac{\sigma^2}{\delta^2}\frac{\partial B_i}{\partial\sigma^2} + W'_\Phi(B_i)\frac{\partial B_i}{\partial\sigma^2}\right)\frac{\partial B_i}{\partial\sigma^2}\right.\end{aligned}$$

$$+\left(-\frac{\sigma^2}{\delta^2}B_i+W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\sigma^4}\Bigg],$$

$$\begin{aligned}\frac{\partial^2\ell_p(\boldsymbol{\theta})}{\partial\delta\partial\sigma^2}&=\frac{n\delta}{(\sigma^2+\delta^2)^2}+\sum_{i=1}^n\left(\frac{1}{\delta^3}B_i^2-\frac{1}{\delta^2}B_i\frac{\partial B_i}{\partial\delta}\right)\\&\quad+\sum_{i=1}^n\Bigg[\left(\frac{2\sigma^2}{\delta^3}B_i-\frac{\sigma^2}{\delta^2}\frac{\partial B_i}{\partial\delta}+W'_\Phi(B_i)\frac{\partial B_i}{\partial\delta}\right)\frac{\partial B_i}{\partial\sigma^2}\\&\quad+\left(-\frac{\sigma^2}{\delta^2}B_i+W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\delta\partial\sigma^2}\Bigg],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\ell_p(\boldsymbol{\theta})}{\partial\delta^2}&=-\frac{n(\sigma^2-\delta^2)}{(\sigma^2+\delta^2)^2}+\frac{\sigma^2}{\delta^3}\sum_{i=1}^n\left(-\frac{3}{\delta}B_i^2+2B_i\frac{\partial B_i}{\partial\delta}\right)\\&\quad+\sum_{i=1}^n\Bigg[\left(\frac{2\sigma^2}{\delta^3}B_i-\frac{\sigma^2}{\delta^2}\frac{\partial B_i}{\partial\delta}+W'_\Phi(B_i)\frac{\partial B_i}{\partial\delta}\right)\frac{\partial B_i}{\partial\delta}\\&\quad+\left(-\frac{\sigma^2}{\delta^2}B_i+W_\Phi(B_i)\right)\frac{\partial^2 B_i}{\partial\delta^2}\Bigg],\end{aligned}$$

where $W'_\Phi(x)=-W_\Phi(x)(x+W_\Phi(x))$. The first and second derivatives of B_i, $i=1,\ldots,n$, in relation to $\boldsymbol{\theta}$ are given by

$$\begin{aligned}\frac{\partial B_i}{\partial\boldsymbol{\beta}}&=-\frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}(\mathbf{x}_i-\rho\mathbf{x}_{i-1}),\\\frac{\partial B_i}{\partial\boldsymbol{\gamma}}&=-\frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}(\mathbf{n}_i-\rho\mathbf{n}_{i-1}),\\\frac{\partial B_i}{\partial\rho}&=-\frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}(y_{i-1}-\mu_{i-1}),\\\frac{\partial B_i}{\partial\sigma^2}&=-\frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3(\sigma^2+\delta^2)^{3/2}}(y_i-\xi_i+b\delta),\\\frac{\partial B_i}{\partial\delta}&=\frac{\sigma^2(y_i-\xi_i+b\delta)+b\delta(\sigma^2+\delta^2)}{\sigma(\sigma^2+\delta^2)^{3/2}}\\\frac{\partial^2 B_i}{\partial\boldsymbol{\beta}\partial\beta^\top}&=\frac{\partial^2 B_i}{\partial\boldsymbol{\gamma}\partial\beta^\top}=\frac{\partial^2 B_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}=\mathbf{0},\\\frac{\partial^2 B_i}{\partial\rho^2}&=0,\\\frac{\partial^2 B_i}{\partial\rho\partial\boldsymbol{\beta}}&=\frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}\mathbf{x}_{i-1},\\\frac{\partial^2 B_i}{\partial\rho\partial\boldsymbol{\gamma}}&=\frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}\mathbf{n}_{i-1}\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 B_i}{\partial\sigma^2\partial\boldsymbol{\beta}} &= \frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3(\sigma^2+\delta^2)^{3/2}}(\mathbf{x}_i-\rho\mathbf{x}_{i-1}),\\
\frac{\partial^2 B_i}{\partial\sigma^2\partial\boldsymbol{\gamma}} &= \frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3(\sigma^2+\delta^2)^{3/2}}(\mathbf{n}_i-\rho\mathbf{n}_{i-1}),\\
\frac{\partial^2 B_i}{\partial\sigma^2\partial\rho} &= \frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3(\sigma^2+\delta^2)^{3/2}}(y_{i-1}-\mu_{i-1}),\\
\frac{\partial^2 B_i}{\partial\delta\partial\boldsymbol{\beta}} &= -\frac{\sigma}{(\sigma^2+\delta^2)^{3/2}}(\mathbf{x}_i-\rho\mathbf{x}_{i-1}),\\
\frac{\partial^2 B_i}{\partial\delta\partial\boldsymbol{\gamma}} &= -\frac{\sigma}{(\sigma^2+\delta^2)^{3/2}}(\mathbf{n}_i-\rho\mathbf{n}_{i-1}),\\
\frac{\partial^2 B_i}{\partial\delta\partial\rho} &= -\frac{\sigma}{(\sigma^2+\delta^2)^{3/2}}(y_{i-1}-\mu_{i-1}),\\
\frac{\partial^2 B_i}{\partial\sigma^4} &= -\frac{\delta(y_i-\xi_i+b\delta)}{4\sigma^5(\sigma^2+\delta^2)^{5/2}}[4\sigma^2(\sigma^2+\delta^2)-3(2\sigma^2+\delta^2)^2],\\
\frac{\partial^2 B_i}{\partial\sigma^2\partial\delta} &= \frac{\sigma(\sigma^2+\delta^2)(y_i-\xi_i+2b\delta)-\frac{1}{2}\left[\sigma^2(y_i-\xi_i+b\delta)+b\delta(\sigma^2+\delta^2)\right]\left(\frac{\sigma^2+\delta^2}{\sigma}+3\sigma\right)}{\sigma^2(\sigma^2+\delta^2)^{5/2}},\\
\frac{\partial^2 B_i}{\partial\delta^2} &= \frac{b(\sigma^2+\delta^2)(2\sigma^2+3\delta^2)-3\delta\left[\sigma^2(y_i-\xi_i+b\delta)+b\delta(\sigma^2+\delta^2)\right]}{\sigma(\sigma^2+\delta^2)^{5/2}},
\end{aligned}
$$

where $\mathbf{x}_0=\mathbf{n}_0=\mathbf{0}$ and $y_0=\mu_0=0$.

Appendix B: Derivation of the matrices N and K

Let $\mathbf{t}=(t_1,\ldots,t_m)$ and *ndx* the number of equidistant Knots desired by the user. It follows the commands in R (R Core Team 2019) for construction of the matrices **N** and **K**:

```
require(splines)
bspline=function(x,ndx,bdeg){
xl=min(x)
xr=max(x)
dx=(xr-xl)/ndx
knots=seq(xl-(bdeg-1)*dx,xr+(bdeg-1)*dx,by=dx)
B=splineDesign(knots,x,bdeg+1,0*x,outer.ok=T)
}

B=bspline(t,ndx,3)
D = diag(ncol(B))
for (k in 1:2) D = diff(D)
```

$$\mathbf{N}=bspline(\mathbf{t},ndx,3),\ \mathbf{K}=\mathbf{D}^\top\mathbf{D}.$$

References

Atkinson A (1981) Two graphical displays for outlying and influential observations in regression. Biometrika 68:13

Azzalini A (1985) A class of distributions which includes the normal ones. Scand J Stat 12:171–178

Bondon P (2009) Estimation of autoregressive models with epsilon-skew-normal innovations. J Multivar Anal 100:1761–1776

Breiman L, Friedman JH (1985) Estimating optimal transformations for multiple regression and correlation (with discussion). J Am Stat Assoc 80(391):580–619

Cao CZ, Lin JG, Zhu LX (2010) Heteroscedasticity and/or autocorrelation diagnostics in nonlinear models with ar(1) and symmetrical errors. Stat Pap 51:813–836

Cook RD (1986) Assessment of local influence. J R Stat Soc Ser B 48:133–169

Cook RD, Weisberg S (1982) Residuals and influence in regression. Chapman and Hall/CRC, Boca Raton

de Boor C (2001) A practical guide to spline, revised. Springer, Berlin

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B 39:1–38

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

Eilers PHC, Marx BD (1996) Flexible smoothing with B-splines and penalties. Stat Sci 11(2):89–121

Ferreira CS, Bolfarine H, Lachos VH (2011) Skew scale mixtures of normal distributions: properties and estimation. Stat Methodol 8:154–171

Ferreira CS, Paula GA (2017) Estimation and diagnostic for skew-normal partially linear models. J Appl Stat 44(16):3033–3053

Ferreira G, Castro LM, Lachos VH, Dias R (2013) Bayesian modeling of autoregressive partial linear models with scale mixture of normal errors. J Appl Stat 40(8):1796–1816

Ferreira JT, Steel MFJ (2006) A constructive representation of univariate skewed distributions. J Am Stat Assoc 101:823–829

Genton MG (2004) Skew-elliptical distributions and their applications: a journey beyond normality. Chapman and Hall, London

Green PJ (1987) Penalized likelihood for general semi-parametric regression models. Int Stat Rev 55:245–259

Green PJ (1990) On use of the EM algorithm for penalized likelihood estimation. J R Stat Soc Ser B 52:443–452

Greene WH (2012) Econometric analysis. 7th edn. Prentice Hall, Upper Saddle River

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

Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, vol 1. Wiley, New York

Lee SY, Xu L (2004) Influence analysis of nonlinear mixed-effects models. Comput Stat Data Anal 45:321–341

Liu S (2004) On diagnostics in conditionally heteroscedastic time series models under elliptical distributions. J Appl Probab 41A:393–405

Lu B, Song X (2006) Local influence of multivariate Probit latent variable models. J Multivar Anal 97:1783–1798

Nadarajah S, Kotz S (2003) Skewed distributions generated by the normal kernel. Stat Probab Lett 65:269–277

Paula GA, Medeiros MJ, Vilca-Labra FE (2009) Influence diagnostics for linear models with first-order autoregressive elliptical errors. Stat Probab Lett 79:339–346

R Core Team (2019) R: A language and environment for statistical computing. In: R foundation for statistical computing, Vienna, Austria. http://www.R-project.org/

Relvas CEM, Paula GA (2016) Partially linear models with first-order autoregressive symmetric errors. Stat Pap 57:795–825

Sahu SK, Dey DK, Branco MD (2003) A new class of multivariate distributions with applications to Bayesian regression models. Can J Stat 31:129–150

Segal MR, Bacchetti P, Jewell NP (1994) Variances for maximum penalized likelihood estimates obtained via the EM algorithm. J R Stat Soc Ser B 56:345–352

Sharafi M, Nematollahi AR (2016) Ar(1) model with skew-normal innovations. Metrika 79:1011–1029
Wood S (2017) Generalized additive models, an introduction with R, 2nd edn. Chapman and Hall, London
Zhu H, Lee S (2001) Local influence for incomplete-data models. J R Stat Soc Ser B 63:111–126