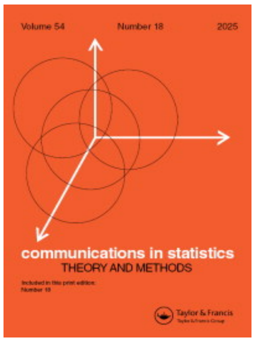


Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: www.tandfonline.com/journals/lsta20

Additive models with *p*-order autoregressive skew-normal errors for modeling trend and seasonality in time series

Clécio S. Ferreira, Gilberto A. Paula & Rodrigo A. Oliveira

To cite this article: Clécio S. Ferreira, Gilberto A. Paula & Rodrigo A. Oliveira (2025) Additive models with *p*-order autoregressive skew-normal errors for modeling trend and seasonality in time series, Communications in Statistics - Theory and Methods, 54:18, 5703-5725, DOI: 10.1080/03610926.2024.2444519

To link to this article: https://doi.org/10.1080/03610926.2024.2444519

Published online: 10 Jan 2025.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2025, VOL. 54, NO. 18, 5703–5725
https://doi.org/10.1080/03610926.2024.2444519

Additive models with *p*-order autoregressive skew-normal errors for modeling trend and seasonality in time series

Clécio S. Ferreira[a], Gilberto A. Paula[b], and Rodrigo A. Oliveira[c]

[a]Department of Statistics, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil; [b]Institute of Mathematics and Statistics, University of São Paulo, São Paulo, SP, Brazil; [c]Institute of Mathematics and Statistics, University of São Paulo, São Paulo, SP, Brazil

ABSTRACT

In this article, we propose an additive model in which the random error follows a skew-normal *p*-order autoregressive (AR) process where the systematic component is approximated by cubic and cyclic cubic regression splines. The maximum likelihood estimators are calculated through the expectation-maximization (EM) algorithm with analytic expressions for the E and M-steps. The effective degrees of freedom concerning the non parametric component are estimated based on a linear smoother. The smoothing parameters are estimated by minimizing the Bayesian information criterion. The conditional quantile residuals are used to construct simulated confidence bands for assessing departures from the error assumptions. Also, we use the same residuals to construct graphs of the autocorrelation and partial autocorrelation functions to verify the AR structure's adequacy for the errors. We then perform local influence analysis based on the conditional expectation of the complete-data log-likelihood function. A simulation study is carried out to evaluate the efficiency of the EM algorithm. Finally, the method is illustrated by using a real dataset of the average weekly cardiovascular mortality in Los Angeles.

ARTICLE HISTORY
15 December 2023
13 December 2024

KEYWORDS
Autoregressive models; additive models; EM algorithm; penalized smoothing; skew-normal distribution

1. Introduction

Additive models (Hastie and Tibshirani 1990) are an important tool when the relationship between explanatory variables and the response emerges in the model in an unspecified association form. In general, its formulation is given by $y = \sum_{j=1}^{q} f_j(t_j) + \epsilon$, where y is the response variable, $t_1, \ldots, t_q$ are the explanatory variables, f_j are smooth arbitrary functions, and ϵ is the random error. In this context, Racine, Su, and Ullah (2014, Chap. 5) proposed a model to explain the growth rate of gross domestic product among countries; Faraway (2016, Chap. 15) predicted atmospheric ozone concentration; Harezlak, Ruppert, and Wand (2018, Chap. 3) used additive models to study test scores of children in California school districts; and Yu and Ruppert (2002) evaluated how the concentration of the air pollutant ozone depends on three meteorological variables; wind speed, temperature, and radiation.

When the hypothesis of independence of the random errors is unrealistic, for example in time series data, we can use some structure of dependence, such as autoregressive processes.

CONTACT Clécio S. Ferreira clecio.ferreira@ufjf.br Department of Statistics, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil.

For instance, in the process of fitting the relationship between temperature and electricity usage, Engle, Granger, and Weiss (1986) modeled the data by first-order autoregressive (AR(1)) errors; Liu (2004) derived some procedures in conditional heteroscedastic time series models under elliptical errors; Paula, Medeiros, and Vilca-Labra (2009) focused on diagnostic methods in linear models with AR(1) elliptical errors; Cao, Lin, and Zhu (2010) assessed heteroscedasticity and autocorrelation in non linear models with AR(1) symmetric errors; and Ferreira et al. (2013) developed a Bayesian analysis for partially linear models (PLM) with first-order autoregressive errors, assumed to belong to the class of the scale mixtures of normal distributions. In the context of semiparametric time series models with autocorrelated errors, Relvas and Paula (2016) derived an iterative Fisher scoring procedure as well as some diagnostic techniques in partially linear models with AR(1) conditional symmetric errors; Zheng and Li (2018) investigated the regression coefficient and autoregressive order shrinkage and selection via the smoothly clipped absolute deviation penalty for a PLM with p-order autoregressive and a divergent number of covariates; and more recently Oliveira and Paula (2021) proposed additive models with p-order autoregressive (AR(p)) conditional symmetric errors based on penalized regression splines for modeling trend and seasonality in time series by cubic and cyclic regression splines, respectively.

These models were developed in the context of symmetric random errors. However, in some situations, the data are asymmetric and the methods mentioned above may not be appropriate. In this sense, a distribution that accommodates skewness and includes the normal distribution as a special case was introduced by Sahu, Dey, and Branco (2003), named skew-normal (SN). This class has been studied by various authors in different contexts, (see, e.g., Genton 2004; Bayes and Branco 2007; Harvey et al. 2010; Ferreira, Vilca, and Bolfarine 2018, among others).

Therefore, a natural and important extension to be considered is an additive model with p-order autoregressive skew-normal errors. Such an extension is useful to accommodate the inherent variability to the data under varied circumstances. Furthermore, autoregressive structures with higher orders usually can adapt well to other structures of dependency, resulting in good fits. Also, in this article we develop the EM (expectation-maximization) algorithm with closed-form expressions for all the estimators of the parameters of the proposed model.

This article is structured as follows. Section 2 specifies the proposed additive models with p-order autoregressive skew-normal errors. Section 3 presents the observed-data log-likelihood function, the EM algorithm, the effective degrees of freedom, and the residual analysis using quantile residuals. Section 5 describes simulation studies to evaluate the efficiency of the EM algorithm. Section 6 illustrates the application of the proposed method to real data on the average weekly cardiovascular mortality in Los Angeles. Finally, Section 7 summarizes our contributions.

2. Specification of the model

The skew-normal (SN) distribution (Azzalini 1985) is a useful distribution for modeling asymmetric data. There are several other versions of this distribution (see Nadarajah and Kotz 2003; Genton 2004; Ferreira and Steel 2006, among others). Here we use another version of the skew-normal, developed by Sahu, Dey, and Branco (2003), where there is a one-to-one relationship between the parameters of the two versions of the univariate case. An advantage of this version is that it directly allows analytic expressions in the E and M-steps for the EM algorithm in the proposed model.

A random variable Y follows a univariate skew-normal distribution (Sahu, Dey, and Branco 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 $N(0,1)$, namely $Y \sim SN(\mu,\sigma^2,\delta)$. For $\delta=0$, the pdf in (1) reduces to that of the normal distribution. The mean and variance of Y are, respectively, given by:

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

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 then 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 additive model with AR(p) skew-normal errors, called ASN-AR(p), is defined as follows:

$$\begin{aligned} y_i &= \mu_i + \epsilon_i, \\ \epsilon_i &= \psi_1\epsilon_{i-1} + \cdots + \psi_p\epsilon_{i-p} + e_i, \text{and} \\ e_i &\sim SN(-b\delta,\sigma^2,\delta), \quad i=1,\ldots,n, \end{aligned} \tag{4}$$

where y_i's denote the response values, $\mu_i = f_T(t_i) + f_S(s_i)$ is the expected value of Y_i, $f_T(\cdot)$ and $f_S(\cdot)$ are smoothing functions, $\psi_1,\cdots,\psi_p$ are the autoregressive parameters, t_i and s_i denote, respectively, the ith time and the time of the year respective to the ith time, and e_i's are independent random errors with a skew-normal distribution with zero mean for $i=1,\ldots,n$. We assume $\epsilon_0 = \epsilon_{-1} = \cdots = \epsilon_{-p+1} = 0$. Let $\psi(B) = 1 - \psi_1 B - \ldots - \psi_p B^p$, considered as a polynomial in B of degree p and $Z_i = y_i - \mu_i$, $i=1,\ldots,n$. So, $Z_i = \sum_{j=1}^{p}\psi_j Z_{i-j} + e_i$ and $\psi(B)Z_i = e_i$ and the process in (4) is said to be *stationary* if all roots of $\psi(B)=0$ are greater than 1 in absolute value (Box, Jenkins, and Reinsel 2008, Sec. 1.2.1).

In addition, we assume that the smoothing functions $f_T(t)$ and $f_S(s)$ are approximated by cubic and cyclic cubic regression splines with fixed knots at the values t_0^j $(j=1,\ldots,r_T)$ and s_0^l $(l=1,\ldots,r_S)$, respectively. There are many equivalent bases that can be used to represent cubic and cyclic splines. In the following, we consider a penalized cubic regression spline and its cyclic version as developed by Wood (2017). The assumptions are that the cubic regression spline must be continuous to the second derivative at the knots t_0^j $(j=1,\ldots,r_T)$ and must have zero second derivative at the first and the last knots. Consequently, it may be written in the linear form

$$f_T(t) = \sum_{j=1}^{r_T} n_{T_j}(t)\gamma_{T_j},$$

where $n_{T_j}(t)$ denotes the basis functions and γ_{T_j} are unknown parameters, for $j = 1, ..., r_T$. Similarly, we write:

$$f_S(s) = \sum_{j=1}^{r_S - 1} n_{S_j}(s)\gamma_{S_j}.$$

Hence, the systematic component of the model (3) may be expressed in the matrix form $\boldsymbol{\mu} = \mathbf{N}_T\boldsymbol{\gamma}_T + \mathbf{N}_S\boldsymbol{\gamma}_S$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, $\mathbf{N}_T$ denotes an $n \times r_T$ matrix with rows $\mathbf{n}_{T_i}^\top = (N_{T_1}(t_i), \ldots, N_{T_{r_T}}(t_i))$, for $i = 1, \ldots, n$ and $\boldsymbol{\gamma}_T = (\gamma_1, ..., \gamma_{r_T})^\top$, $\mathbf{N}_S$ is an $n \times (r_S - 1)$ matrix with rows $\mathbf{n}_{S_i}^\top = (N_{S_1}(t_i), \ldots, N_{S_{r_S-1}}(t_i))$, for $i = 1, \ldots, n$ and $\boldsymbol{\gamma}_S = (\gamma_1, ..., \gamma_{r_S-1})^\top$. In order to obtain a smooth fit, we consider a continuous penalization of the likelihood function from which an EM iterative process can be developed. However, discrete penalization has been widely applied with B-splines, such as P-splines (Eilers and Marx 1996).

Some authors have discussed the problems of identifiability of additive models. For example, Lin and Kulasekera (2007) provided a proof of the identifiability for both single-index models and partially linear single index models assuming only the continuity of the regression function. Yuan (2011) imposed some restrictions based on linear ridge function, quadratic ridge function, and projection indices matrix of column full rank. Stringer (2024) used linear constraints of the type $\mathbf{c}_j^\top \boldsymbol{\gamma}_j = 0$, for $\mathbf{c}_j \in \mathbf{R}^{r_j}$ to impose identifiability in generalized additive models. Finally, Wood (2017, Sec. 4.3) proposed the restriction $\sum_{i=1}^n f_1(x_i) = 0$ to deal with the identifiability problem. This is the form we use in this article, where the matrices $\mathbf{N}_T$ and $\mathbf{N}_S$ are constructed using the function "smoothCon" of the package *mgcv* in R (R Core Team 2023). So, it can be shown that these matrices are full column rank and consequently the matrices AN_T and AN_S are also, which guarantees the identification of the additive functions.

3. Statistical inference

In this section, we discuss some inferential aspects of the ASN-AR(p) model as well as the penalized maximum likelihood estimation using the EM algorithm. In turn, the EM algorithm (Dempster, Laird, and Rubin 1977) used to obtain the maximum likelihood estimate of $\boldsymbol{\theta}$ and a discussion of degrees of freedom estimation are given in Section 3.1; the standard error estimation of $\widehat{\boldsymbol{\theta}}$ is presented in Appendix A; and the residual analysis based on conditional quantile residuals is discussed in Section 3.2.

From (4) the observed-data log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\gamma}_T^\top, \boldsymbol{\gamma}_S^\top, \sigma^2, \delta, \boldsymbol{\psi}^\top)^\top \in \mathbb{R}^{p^*}$, where $\boldsymbol{\psi} = (\psi_1, \cdots, \psi_p)^\top$ and $p^* = r_T + r_S + p$, can be expressed as:

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

where $B_i = \frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}(y_i - \xi_i + b\delta)$, $\xi_i = \mu_i + \sum_{j=1}^p \psi_j(y_{i-j} - \mu_{i-j})$, $\mu_i = \mathbf{n}_{T_i}^\top\boldsymbol{\gamma}_T + \mathbf{n}_{S_i}^\top\boldsymbol{\gamma}_S$, for $i = 1, \ldots, n$.

There are two possible difficulties in maximizing $\ell(\boldsymbol{\theta})$ in (5), namely the evaluation of the term $\Phi(\cdot)$ and the need of to impose some restriction on $f_T(t)$ and $f_S(s)$ to avoid overfitting and non identification of $\boldsymbol{\gamma}_T$ and $\boldsymbol{\gamma}_S$, respectively (see, for instance, Green 1987). So, we propose the parameter estimation evaluated by an EM algorithm of the penalized log-likelihood function, given by:

$$\ell_p(\boldsymbol{\theta},\boldsymbol{\alpha}) = \ell(\boldsymbol{\theta}) - \frac{\alpha_T}{2}J(\boldsymbol{\gamma}_T) - \frac{\alpha_S}{2}J(\boldsymbol{\gamma}_S), \tag{6}$$

where $\boldsymbol{\alpha} = (\alpha_T, \alpha_S)$, and $\alpha_T > 0, \alpha_T > 0$ are the smoothing parameters that will be estimated separately while $J(\boldsymbol{\gamma}_T)$ and $J(\boldsymbol{\gamma}_S)$ denote the penalty functions, defined as follows:

$$J(\boldsymbol{\gamma}_k) = \int_{a_k}^{b_k} [f_k^{(2)}(t)]^2 dt = \boldsymbol{\gamma}_k^\top \mathbf{K}_k \boldsymbol{\gamma}_k, \; k = \{T, S\},$$

(see also Wood (2017), Chap. 5) with $\mathbf{K}_T \in \mathbb{R}^{r_T \times r_T}$ and $\mathbf{K}_S \in \mathbb{R}^{(r_S-1)\times(r_S-1)}$ being non negative definite matrices that depend only on the knot differences.

3.1. Parameter estimation via EM algorithm

The EM algorithm (Dempster, Laird, and Rubin 1977) has been widely applied for to find maximum likelihood estimates. One advantage of the EM algorithm is that the M-step involves only complete data maximum likelihood 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, Kotz, and Balakrishnan 1994). Using the stochastic representation of skew-normal distribution, we have that:

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

After some algebraic manipulations, the joint distribution for (Y_i, Z_i) is 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 the properties of truncated normal distributions, we obtain:

$$\begin{aligned} E[Z_i|y_i] &= \mu_{iz} + \sigma_z W_\Phi(\mu_{iz}/\sigma_z) \text{ and} \\ E[Z_i^2|y_i] &= \mu_{iz}^2 + \sigma_z^2 + \sigma_z \mu_{iz} W_\Phi(\mu_{iz}/\sigma_z), \end{aligned} \tag{7}$$

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 $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{z}^\top)^\top$ can be expressed as

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = C &- \frac{n}{2}\log\sigma^2 \\ &- \frac{1}{2\sigma^2}\sum_{i=1}^n \left[(y_i - \xi_i)^2 - 2\delta(y_i - \xi_i)(z_i - b) + \delta^2(b^2 - 2bz_i + z_i^2)\right], \end{aligned} \tag{8}$$

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

Like in the original proposal of Dempster, Laird, and Rubin (1977), the E-step of our algorithm consists of taking the conditional expectation $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}]$, where $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\gamma}}_T^{(k)\top}, \widehat{\boldsymbol{\gamma}}_S^{(k)\top}, \widehat{\sigma^2}^{(k)}, \widehat{\delta}^{(k)}, \widehat{\boldsymbol{\psi}}^{(k)\top})^\top$, is the current estimate of $\boldsymbol{\theta}$ at the kth iteration.

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

$$Q_p(\widehat{\boldsymbol{\theta}}^{(k)}) = Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) - \frac{\alpha_T}{2}\boldsymbol{\gamma}_T^\top \mathbf{K}_T \boldsymbol{\gamma}_T - \frac{\alpha_S}{2}\boldsymbol{\gamma}_S^\top \mathbf{K}_S \boldsymbol{\gamma}_S. \quad (9)$$

Given $\widehat{\boldsymbol{\alpha}}^{(k)} = (\widehat{\boldsymbol{\alpha}}_T^{(k)}, \widehat{\boldsymbol{\alpha}}_S^{(k)})$, the M-step consists maximizing $Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$. After some simple algebra, that the conditional expectation of the complete log-likelihood function has the form:

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= C - \frac{n}{2}\log\sigma^2 \quad (10)\\ &- \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(y_i-\xi_i)^2 - 2\delta(y_i-\xi_i)(\widehat{z}_i^{(k)} - b) + \delta^2(b^2 - 2b\widehat{z}_i^{(k)} + \widehat{z^2}_i^{(k)})\right], \end{aligned}$$

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

3.1.1. Step-by-step instructions for the EM algorithm

Let $\mathbf{A} = \mathbf{A}(\boldsymbol{\psi})$ be an $(n \times n)$ matrix given by:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\psi_1 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\psi_2 & -\psi_1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -\psi_p & -\psi_{p-1} & -\psi_{p-2} & \cdots & -\psi_1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & -\psi_p & -\psi_{p-1} & \cdots & -\psi_2 & -\psi_1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & -\psi_2 & -\psi_1 & 1 \end{pmatrix}.$$

The EM algorithm for the ASN-AR(p) model can be summarized in the following steps:

(1) *E-step*: Given the current estimates $\widehat{\boldsymbol{\theta}}^{(k)}$ and $\widehat{\boldsymbol{\alpha}}^{(k)}$ at the kth iteration and using (7), we obtain the conditional expectation of the complete data log-likelihood function given the observed $\mathbf{y}$, called the Q-function, which is given by (3.1), such that:

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

and

$$\widehat{z^2}_i^{(k)} = [\widehat{\mu}_{iz}^{(k)}]^2 + \widehat{\sigma^2}_z^{(k)} + \widehat{\sigma}_z^{(k)}\widehat{\mu}_{iz}^{(k)} W_\Phi\left(\frac{\widehat{\mu}_{iz}^{(k)}}{\widehat{\sigma}_z^{(k)}}\right), \quad (12)$$

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 $\widehat{\xi}_i^{(k)} = \widehat{\mu}_i^{(k)} + \sum_{j=1}^{p}\widehat{\psi}_j^{(k)}(y_{i-j} - \widehat{\mu}_{i-j}^{(k)})$, with $\widehat{\mu}_i^{(k)} = \mathbf{n}_{T_i}^\top\widehat{\boldsymbol{\gamma}}_T^{(k)} + \mathbf{n}_{S_i}^\top\widehat{\boldsymbol{\gamma}}_S^{(k)}$, $i = 1, \ldots, n$.

(2) *M-step* : Fix $\widehat{\alpha}_T^{(k)}$ and $\widehat{\alpha}_S^{(k)}$, update $\widehat{\boldsymbol{\gamma}}_T^{(k)}$, $\widehat{\boldsymbol{\gamma}}_S^{(k)}$, $\widehat{\sigma^2}^{(k)}$, $\widehat{\delta}^{(k)}$, and $\widehat{\boldsymbol{\psi}}^{(k)}$ as

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

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

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

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

$$\widehat{\psi}_j^{(k+1)} = \frac{1}{\sum_{i=1}^{n} r_{i-j}^2}\sum_{i=1}^{n}\left[r_i - \sum_{\substack{l=1\\ l\neq j}}^{p}\widehat{\psi}_l^{(k)} r_{i-l} - \widehat{\delta}^{(k)}(\widehat{z}_i^{(k)} - b)\right] r_{i-j}, \quad j = 1, \cdots, p,$$

where

$$\widehat{\mathbf{S}}_T^{(k)}(\alpha_T) = \left[(\widehat{\mathbf{A}}^{(k)}\mathbf{N}_T)^\top\widehat{\mathbf{A}}^{(k)}\mathbf{N}_T + \alpha_T\widehat{\sigma^2}^{(k)}\mathbf{K}_T\right]^{-1}(\widehat{\mathbf{A}}^{(k)}\mathbf{N}_T)^\top \quad \text{and}$$

$$\widehat{\mathbf{S}}_S^{(k)}(\alpha_S) = \left[(\widehat{\mathbf{A}}^{(k)}\mathbf{N}_S)^\top\widehat{\mathbf{A}}^{(k)}\mathbf{N}_S + \alpha_S\widehat{\sigma^2}^{(k)}\mathbf{K}_S\right]^{-1}(\widehat{\mathbf{A}}^{(k)}\mathbf{N}_S)^\top$$

are smoothers with $\widehat{\mathbf{y}}^{*(k)} = \widehat{\mathbf{A}}^{(k)}\mathbf{y} - \widehat{\delta}^{(k)}(\widehat{\mathbf{z}}^{(k)} - b)$ being a pseudo-response, $r_i = y_i - \widehat{\mu}_i^{(k)} = y_i - \mathbf{n}_{T_i}^\top\widehat{\boldsymbol{\gamma}}_T^{(k)} - \mathbf{n}_{S_i}^\top\widehat{\boldsymbol{\gamma}}_S^{(k)}$, $i = 1, ..., n$, $r_0 = r_{-1} = \cdots = r_{-(p-1)} = 0$, and $\widehat{\mathbf{z}}^{(k)} = (\widehat{z}_1^{(k)}, ..., \widehat{z}_n^{(k)})^\top$, for $k = 0, 1, 2, \ldots$.

(3) Given $\widehat{\boldsymbol{\theta}}^{(k+1)}$, update $\widehat{\boldsymbol{\alpha}}^{(k)}$ by minimizing the Bayesian information criterion (BIC),

$$\widehat{\boldsymbol{\alpha}}^{(k+1)} = \text{argmin}_{\boldsymbol{\alpha}}\text{BIC}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k+1)}),$$

given by

$$\text{BIC}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}) = -2\ell_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha}) + p(\boldsymbol{\alpha})\log n,$$

where $\ell_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha})$ denotes the penalized log-likelihood function available at $\widehat{\boldsymbol{\theta}}$ for a fixed $\boldsymbol{\alpha}$, as defined in (6).

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

$$\widehat{\boldsymbol{\gamma}}_T^{(k+1)} = \left[(\widehat{\mathbf{A}}^{(k)}\mathbf{N}_T)^\top\widehat{\mathbf{W}}_T^{(k)}\widehat{\mathbf{A}}^{(k)}\mathbf{N}_T + \alpha_T\widehat{\sigma^2}^{(k)}\mathbf{K}_T\right]^{-1}(\widehat{\mathbf{A}}^{(k)}\mathbf{N}_T)^\top\widehat{\mathbf{W}}_T^{(k)}\widehat{\mathbf{y}}^{*(k)} \quad \text{and}$$

$$\widehat{\boldsymbol{\gamma}}_S^{(k+1)} = \left[(\widehat{\mathbf{A}}^{(k)}\mathbf{N}_S)^\top\widehat{\mathbf{W}}_S^{(k)}\widehat{\mathbf{A}}^{(k)}\mathbf{N}_S + \alpha_S\widehat{\sigma^2}^{(k)}\mathbf{K}_S\right]^{-1}(\widehat{\mathbf{A}}^{(k)}\mathbf{N}_S)^\top\widehat{\mathbf{W}}_S^{(k)}\widehat{\mathbf{y}}^{*(k)},$$

where

$$\mathbf{W}_T = \mathbf{I}_n - \mathbf{A}\mathbf{N}_T[(\mathbf{A}\mathbf{N}_T)^\top\mathbf{A}\mathbf{N}_T + \alpha_T\sigma^2\mathbf{K}_T]^{-1}(\mathbf{A}\mathbf{N}_T)^\top \quad \text{and}$$

$$\mathbf{W}_S = \mathbf{I}_n - \mathbf{A}\mathbf{N}_S[(\mathbf{A}\mathbf{N}_S)^\top\mathbf{A}\mathbf{N}_S + \alpha_S\sigma^2\mathbf{K}_S]^{-1}(\mathbf{A}\mathbf{N}_S)^\top.$$

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

$$\widehat{\mathbf{A}}\mathbf{N}_T\widehat{\boldsymbol{\gamma}}_T = \widehat{\mathbf{H}}_T(\alpha_T)\widehat{\mathbf{y}}^* \quad \text{and} \quad \widehat{\mathbf{A}}\mathbf{N}_S\widehat{\boldsymbol{\gamma}}_S = \widehat{\mathbf{H}}_S(\alpha_S)\widehat{\mathbf{y}}^*,$$

where

$$\widehat{\mathbf{H}}_T(\alpha_T) = \widehat{\mathbf{A}}\mathbf{N}_T[(\widehat{\mathbf{A}}\mathbf{N}_T)^\top \widehat{\mathbf{W}}_T\widehat{\mathbf{A}}\mathbf{N}_T + \alpha_T\widehat{\sigma^2}\mathbf{K}_T]^{-1}(\widehat{\mathbf{A}}\mathbf{N}_T)^\top\widehat{\mathbf{W}}_T,$$
$$\widehat{\mathbf{H}}_S(\alpha_S) = \widehat{\mathbf{A}}\mathbf{N}_S[(\widehat{\mathbf{A}}\mathbf{N}_S)^\top \widehat{\mathbf{W}}_S\widehat{\mathbf{A}}\mathbf{N}_S + \alpha_S\widehat{\sigma^2}\mathbf{K}_S]^{-1}(\widehat{\mathbf{A}}\mathbf{N}_S)^\top\widehat{\mathbf{W}}_S$$

and so

$$\mathbf{H}(\boldsymbol{\alpha}) = \text{blockdiag}\left(\widehat{\mathbf{H}}_T(\alpha_T), \widehat{\mathbf{H}}_S(\alpha_S)\right)$$

can be interpreted as a linear smoother. As pointed out by (Hastie and Tibshirani 1990, p. 52), the sum of the eigenvalues of $\widehat{\mathbf{H}}(\boldsymbol{\alpha})$, for $\boldsymbol{\alpha}$ fixed, namely

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

can be defined as the effective degrees of freedom due to the non parametric fit.

Then, one has a total of $p(\boldsymbol{\alpha}) = p + 2 + df(\boldsymbol{\alpha})$ parameters to be used in the BIC.

Notes on implementation

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 can be obtained by computing $\widehat{\sigma}^{2(0)}$ as the standard deviation of $\mathbf{y}$. So, for given α_T and α_S, $\widehat{\boldsymbol{\gamma}}_T^{(0)} = (\mathbf{N}_T^\top\mathbf{N}_T + \alpha_T\widehat{\sigma}^{2(0)}\mathbf{K}_T)^{-1}\mathbf{N}_T^\top\mathbf{y}$, $\widehat{\boldsymbol{\gamma}}_S^{(0)} = (\mathbf{N}_S^\top\mathbf{N}_S + \alpha_S\widehat{\sigma}^{2(0)}\mathbf{K}_S)^{-1}\mathbf{N}_S^\top\mathbf{y}$, $\widehat{\delta}^{(0)}$ can be the sample skewness coefficient of $\mathbf{y} - \mathbf{N}_T\widehat{\boldsymbol{\gamma}}_T^{(0)} - \mathbf{N}_S\widehat{\boldsymbol{\gamma}}_S^{(0)}$ and $\widehat{\boldsymbol{\psi}}^{(0)} = \mathbf{0}$.

We use the "*optim*" routine in R (R Core Team 2023) to estimate $\boldsymbol{\alpha}$, with $\boldsymbol{\alpha}$ between $(0, 10^3) \times (0, 10^3)$.

3.2. Residual analysis

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

$$r_{q_i} = y_i - \mathbf{n}_{T_i}^\top\widehat{\boldsymbol{\gamma}}_T - \mathbf{n}_{S_i}^\top\widehat{\boldsymbol{\gamma}}_S - \sum_{j=1}^{p}\widehat{\psi}_j\hat{\epsilon}_{i-j} \tag{13}$$
$$= \begin{cases} y_i - \mathbf{n}_{T_i}^\top\widehat{\boldsymbol{\gamma}}_T - \mathbf{n}_{S_i}^\top\widehat{\boldsymbol{\gamma}}_S, & i = 1, ..., p, \\ y_i - \mathbf{n}_{T_i}^\top\widehat{\boldsymbol{\gamma}}_T - \mathbf{n}_{S_i}^\top\widehat{\boldsymbol{\gamma}}_S - \sum_{j=1}^{p}\hat{\psi}_j\left[y_{i-j} - \mathbf{n}_{T_{i-j}}^\top\widehat{\boldsymbol{\gamma}}_T - \mathbf{n}_{S_{i-j}}^\top\widehat{\boldsymbol{\gamma}}_S\right], & i = p+1, ..., n. \end{cases}$$

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

$$t_{q_i} = \Phi^{-1}(F_Y(r_{q_i}; -b\widehat{\delta}, \widehat{\sigma}^2, \widehat{\delta})), \quad i = 1, ..., n. \tag{14}$$

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

4. The local influence approach

The main objective of sensitivity studies is to evaluate changes in the parameter estimates, mainly inferential changes, under perturbations in the model or data. The method proposed by Cook (1986), called local influence, consists of studying the influence of small perturbations in the model or the data on the parameter estimates. The method is developed using maximum likelihood estimation through likelihood displacement, which is not the case here because we use the EM approach. For this case, Zhu and Lee (2001) proposed an extension of the local influence method when the EM algorithm is used for maximum likelihood estimation. The approach is based on the Q-displacement instead of the likelihood displacement. We propose a natural extension of this method to the proposed model.

Consider $\mathbf{w} = (w_1, ..., w_n)^\top$ to be a perturbation vector varying in an open region $\mathbf{\Omega} \in \mathbb{R}^n$, $\ell_{c_p}(\boldsymbol{\theta}, \mathbf{w}|\mathbf{y}_c)$, where $\boldsymbol{\theta} \in \mathbb{R}^p$ is the complete-data penalized log-likelihood function of the perturbed model and $\widehat{\boldsymbol{\theta}}(\mathbf{w})$ denotes the maximum value 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 assumed that $\mathbf{w}_0$ exists such that $\ell_{c_p}(\boldsymbol{\theta}, \mathbf{w}_0|\mathbf{Y}_c) = \ell_{c_p}(\boldsymbol{\theta}|\mathbf{Y}_c)$ for all $\boldsymbol{\theta}$. The influence graph is defined as $\boldsymbol{\alpha}(\mathbf{w}) = (\mathbf{w}^\top, f_Q(\mathbf{w}))^\top$, where the Q-displacement function $f_Q(\mathbf{w})$ is defined as

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

The normal curvature $C_{f_Q,\mathbf{d}}$, of $\boldsymbol{\alpha}(\mathbf{w})$ calculated 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) showed 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 $\Delta_{\mathbf{w}} = \left.\dfrac{\partial^2 Q_p(\boldsymbol{\theta}, \mathbf{w}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\mathbf{w}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\mathbf{w})}$.

A measure of influence is constructed by calculating 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 $\xi_1 \geq \ldots \geq \xi_q, \xi_{q+1} = \ldots = \xi_n = 0$ and $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are elements of the associated orthonormal basis. The aggregate contribution vector of all eigenvectors corresponding to non zero eigenvalues is 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 $\mathbf{e}_k^2 = (e_{k1}^2, \ldots, e_{kn}^2)$.

To verify the influential observations, we inspect the graph of $\{M(0)_l, l = 1, \ldots, n\}$. 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, $\ddot{Q}_\theta(\widehat{\theta})$

To obtain the diagnostic measures of the ASN-AR(p) 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 $\boldsymbol{\theta} = (\boldsymbol{\gamma}_S^\top, \boldsymbol{\gamma}_T^\top, \sigma^2, \delta, \boldsymbol{\psi})^\top$. It follows from ((9) that $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})$ has elements given by:

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma}_T \partial \boldsymbol{\gamma}_T^\top} = -\frac{1}{\sigma^2}(\mathbf{A}\mathbf{N}_T)^\top(\mathbf{A}\mathbf{N}_T) - \alpha_T \mathbf{K}_T,$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma}_S \partial \boldsymbol{\gamma}_S^\top} = -\frac{1}{\sigma^2}(\mathbf{A}\mathbf{N}_S)^\top(\mathbf{A}\mathbf{N}_S) - \alpha_S \mathbf{K}_S,$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma}_S \partial \boldsymbol{\gamma}_T^\top} = -\frac{1}{\sigma^2}(\mathbf{A}\mathbf{N}_S)^\top(\mathbf{A}\mathbf{N}_T),$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \boldsymbol{\gamma}_r} = -\frac{1}{\sigma^4}(\mathbf{A}\mathbf{N}_r)^\top\left[\mathbf{y} - \boldsymbol{\xi} - \delta(\widehat{\mathbf{z}} - b\mathbf{1}_n)\right], \quad \text{for } r \in \{S, T\},$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \delta \partial \boldsymbol{\gamma}_r} = -\frac{1}{\sigma^2}(\mathbf{A}\mathbf{N}_r)^\top(\widehat{\mathbf{z}} - b\mathbf{1}_n), \quad \text{for } r \in \{S, T\},$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\psi}_j \partial \boldsymbol{\gamma}_r} = -\frac{1}{\sigma^2}(\mathbf{A}\mathbf{N}_r)^\top \mathbf{e}_j, \quad \text{for } r \in \{S, T\} \text{ and } j = 1, ..., p,$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \sigma^4} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}\left[(y_i - \xi_i)^2 - 2\delta(y_i - \xi_i)(\widehat{z}_i - b) + \delta^2(b^2 - 2b\widehat{z}_i + \widehat{z^2}_i)\right],$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \delta \partial \sigma^2} = \frac{1}{\sigma^4}\sum_{i=1}^{n}\left[-(y_i - \xi_i)(\widehat{z}_i - b) + \delta(b^2 - 2b\widehat{z}_i + \widehat{z^2}_i)\right],$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \psi_j \partial \sigma^2} = -\frac{1}{\sigma^4}\mathbf{e}_j^\top\left(\mathbf{y} - \boldsymbol{\xi} - \delta(\widehat{\mathbf{z}} - b\mathbf{1}_n)\right), \quad \text{for } j = 1, ..., p,$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \delta^2} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(b^2 - 2b\widehat{z}_i + \widehat{z^2}_i),$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \psi_j \partial \delta} = -\frac{1}{\sigma^2}\mathbf{e}_j^\top(\widehat{\mathbf{z}} - b\mathbf{1}_n), \quad \text{for } j = 1, ..., p, \text{ and}$$

$$\frac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \psi_l \psi_j} = -\frac{1}{\sigma^2}\mathbf{e}_l^\top \mathbf{e}_j, \quad \text{for } l, j = 1, ..., p,$$

where $\mathbf{1}_n$ is an all ones vector, $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)^\top$, $\mathbf{e}_j = \{\mathbf{0}_j\ e_1\ e_2\ ...\ e_{n-j}\}^\top$ with $e_l = y_l - \mu_l$, $l = 1, ..., n - j$, and $\mathbf{0}_j$ is a vector of zeros of length j, $j = 1, ..., p$.

4.2. Perturbation schemes

In this section, we consider the three usual perturbation schemes in local influence for the ASN-AR(p) 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

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

with $Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) \propto -\frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left[(y_i-\xi_i)^2 - 2\delta(y_i-\xi_i)(z_i-b) + \delta^2(b^2-2bz_i+z_i^2)\right]$.

In this case, $\mathbf{w}_0 = \mathbf{1}_n$ and $\dfrac{\partial Q_p(\boldsymbol{\theta},\mathbf{w}|\widehat{\boldsymbol{\theta}})}{\partial \mathbf{w}_i} = Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ and $\boldsymbol{\Delta}_{\mathbf{w}_0} = \left.\dfrac{\partial^2 Q(\boldsymbol{\theta},\mathbf{w}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\mathbf{w}^\top}\right|_{\mathbf{w}=\mathbf{w}_0} = \left(\boldsymbol{\Delta}_{\boldsymbol{\gamma}_S}^\top, \boldsymbol{\Delta}_{\boldsymbol{\gamma}_T}^\top, \boldsymbol{\Delta}_{\sigma^2}^\top, \boldsymbol{\Delta}_{\delta}^\top, \boldsymbol{\Delta}_{\boldsymbol{\psi}}^\top\right)^\top$ has elements given by

$$\boldsymbol{\Delta}_{\boldsymbol{\gamma}_r} = \frac{1}{\sigma^2}(\mathbf{A}\mathbf{N}_r)^\top\mathbf{D}\left(\mathbf{y}-\boldsymbol{\xi}-\delta(\widehat{\mathbf{z}}-b\mathbf{1}_n)\right) - \alpha_r\mathbf{K}_r\boldsymbol{\gamma}_r\mathbf{1}_n^\top, \quad \text{for } r\in\{S,T\},$$
$$\boldsymbol{\Delta}_{\sigma^2} = -\frac{1}{2\sigma^2}\mathbf{1}_n^\top + \frac{1}{2\sigma^4}\left[(\mathbf{y}-\boldsymbol{\xi})^\top\mathbf{D}(\mathbf{y}-\boldsymbol{\xi}) - 2\delta(\mathbf{y}-\boldsymbol{\xi})^\top\mathbf{D}(\widehat{\mathbf{z}}-b) + \delta^2(b^2-2b\widehat{\mathbf{z}}+\widehat{\mathbf{z}^2})^\top\right]$$
$$\boldsymbol{\Delta}_{\delta} = \frac{1}{\sigma^2}\left[(\mathbf{y}-\boldsymbol{\xi})^\top\mathbf{D}(\widehat{\mathbf{z}}-b) - \delta(b^2-2b\widehat{\mathbf{z}}+\widehat{\mathbf{z}^2})^\top\right]$$
$$\boldsymbol{\Delta}_{\psi_j} = \frac{1}{\sigma^2}\mathbf{e}_j^\top\mathbf{D}\left(\mathbf{y}-\boldsymbol{\xi}-\delta(\widehat{\mathbf{z}}-b\mathbf{1}_n)\right), \quad \text{for } j=1,\dots,p.$$

4.2.2. Response variable perturbation

Here we consider an additive perturbation given by:

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

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

$$\begin{aligned} Q_p(\boldsymbol{\theta}, w_i|\widehat{\boldsymbol{\theta}}) \propto & -\frac{1}{2\sigma^2}\left[(y_{iw}-\xi_{iw})^2 - 2\delta(y_{iw}-\xi_{iw})(\widehat{z}_i-b)\right] \\ & -\frac{1}{2\sigma^2}\left[(y_{i+1,w}-\xi_{i+1,w})^2 - 2\delta(y_{i+1,w}-\xi_{i+1,w})(\widehat{z}_{i+1}-b)\right], \end{aligned}$$

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

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

$$\boldsymbol{\Delta}_{\boldsymbol{\gamma}_r} = \frac{S_y}{\sigma^2}(\mathbf{A}\mathbf{N}_r)^\top, \quad \text{for } r\in\{S,T\},$$
$$\boldsymbol{\Delta}_{\sigma^2} = \frac{S_y}{\sigma^4}\left(\mathbf{y}-\boldsymbol{\xi}-\delta\right)^\top$$
$$\boldsymbol{\Delta}_{\delta} = \frac{S_y}{\sigma^2}\mathbf{1}_n^\top$$
$$\boldsymbol{\Delta}_{\psi_j} = \frac{S_y}{\sigma^2}\mathbf{e}_j^\top, \quad \text{for } j=1,\dots,p.$$

5. Simulation study

In order to evaluate the performance of the method proposed in Section 2, Monte Carlo simulation studies are carried out. We focus on an ASN-AR(2) of the form

$$y_i = f_T(t_i) + f_S(s_i) + \epsilon_i, \quad i=1,\dots,n, \tag{15}$$

where $\sigma^2 = 0.01$, $\delta = 0.7$, $\psi_1 = 0.5$ and $\psi_2 = 0.3$. For the non linear covariates t_i and s_i, we consider a sequence of equally spaced points. The first component $f_T(t)$ is represented by

Table 1. Mean and standard deviations (SD) for EM estimates and empirical standard error estimates (SE_{emp}) based on 1000 samples from model (15).

	True	n=200			n=1000			n=1500		
Parameter	Value	Mean	SD	SE_{emp}	Mean	SD	SE_{emp}	Mean	SD	SE_{emp}
σ^2	0.010	0.012	0.035	0.004	0.010	0.008	0.003	0.010	0.002	0.002
δ	0.700	0.665	0.151	0.049	0.697	0.037	0.020	0.699	0.016	0.016
ψ_1	0.500	0.459	0.070	0.092	0.498	0.022	0.022	0.501	0.018	0.018
ψ_2	0.300	0.250	0.068	0.093	0.298	0.021	0.022	0.301	0.017	0.018

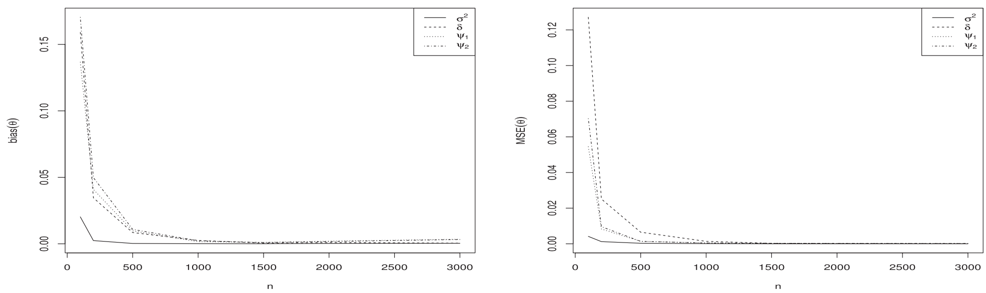


Figure 1. Bias and MSE of the estimates of the parameters σ^2, δ, ψ_1 and ψ_2 for ASN-AR(2).

a function of the form $f_T(t) = 2 - 5 \times t + 5 \times \exp(-100 \times ((t - 0.5)^2))$, $t \in (1/n, 1)$, while the second component $f_S(s)$ is given by $f_S(s) = 2 \times \sin(100 \times s/(n \times \pi))$, $s \in (1, n)$. Also, we use 9 knots for f_T and 5 knots for f_S in grids of equally spaced points for each interval of the data.

The simulation studies are based on $M = 1000$ replications of the above model with different sample sizes, $n = 100, 200, 500, 1000, 1500$, and 3000.

With these estimates, we calculated for θ_k (denoting σ^2, δ, ψ_1 and ψ_2) the *Mean* as $\bar{\widehat{\theta}}_k = \sum_{j=1}^{M} \widehat{\theta}_{k_j}/M$, *SD*=$\sqrt{\sum_{j=1}^{M}(\widehat{\theta}_{k_j} - \bar{\widehat{\theta}}_k)^2/(M-1)}$ and SE_{emp} as $\sqrt{\sum_{j=1}^{M} I_j^{kk}/M}$, where I_j^{kk} denotes the k-th diagonal element of inverse of the observed information matrix.

It can be seen from Table 1 that the *bias* of MLE ($|\bar{\widehat{\theta}}_k - \theta_k|$) decreases when 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 with the sample size, showing that the calculation of the observed information matrix seems to be correct. Also, Figure 1 presents the empirical $bias(\theta_k) = |\theta_k - \bar{\widehat{\theta}}_k|$ and $\text{MSE}(\theta_k) = \frac{1}{M}\sum_{i=1}^{M}(\widehat{\theta}_k - \theta_k)^2 + bias(\theta_k)^2$ for σ^2, δ, ψ_1 and ψ_2. Both bias and MSE tend to 0 when the sample size increases, indicating the convergence of the estimates of the parameters to their true values.

Also, for $n = 3000$, we plotted the histograms of the $M = 1000$ estimates of the parameters σ^2, δ, ψ_1 and ψ_2 for ASN-AR(2), along with the normal densities for mean and the standard deviation of these estimates (Figure 2). The graphs show that the estimates attain the convergence to the normal distribution by the properties of the MLEs. To confirm this information we performed the Shapiro-Wilk test for normality, which presented the following p-values: 0.137 for σ^2, 0.307 for δ, 0.569 for ψ_1 and 0.629 for ψ_2.

Figure 3 presents the results of the simulation study for $M = 1000$ samples from model (15) for $n = 200, 1000$, and 1500. Note that the variability and the bias among the estimates of

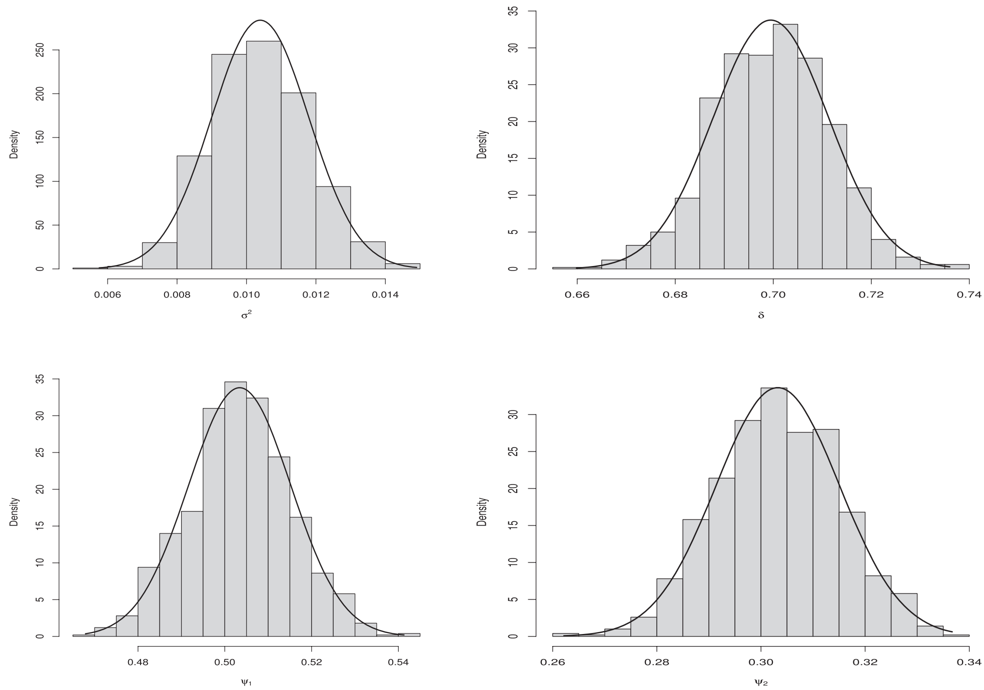


Figure 2. Histograms of the M estimates of the parameters σ^2, δ, ψ_1 and ψ_2 for ASN-AR(2), for $n = 3000$.

the non parametric functions $f_T(t)$ and $f_S(s)$ decrease as the sample size increases, indicating the consistency of the non parametric estimators.

6. Application

To illustrate the method developed in this article, we analyzed the dataset consisting of average weekly cardiovascular mortality in Los Angeles from January 1970 to October 1979, totaling 508 observations (Shumway 1988). The data are available in the R package *astsa*. All daily data are available, that is, there are no missing data.

Figure 4 describes the time series and the empirical distribution, for each month, of the average weekly cardiovascular mortality in Los Angeles. In order to have a kurtosis and skewness flexibility, we fit the new model to understand the trend and seasonality of this time series.

Figure 5 shows the seasonality of the time series over the years, where we can see the high mortalities from November to January. Therefore, it is possible to perceive a persistent seasonality throughout the years, corroborating the existence of a model that presents seasonality.

Under model (16), a residual analysis (based on residual time series plot, PACF and Ljung-Box tests) presented in Figure 6(a-d) indicates autoregressive errors. Therefore, in order to perform a better fitting to the data, we consider the same model given in (16), but now assuming $AR(p)$ errors. Moreover, the standard residuals (Figure 6(a)) also presented signs of skewness, which suggests the use of a skewed distribution to model the errors.

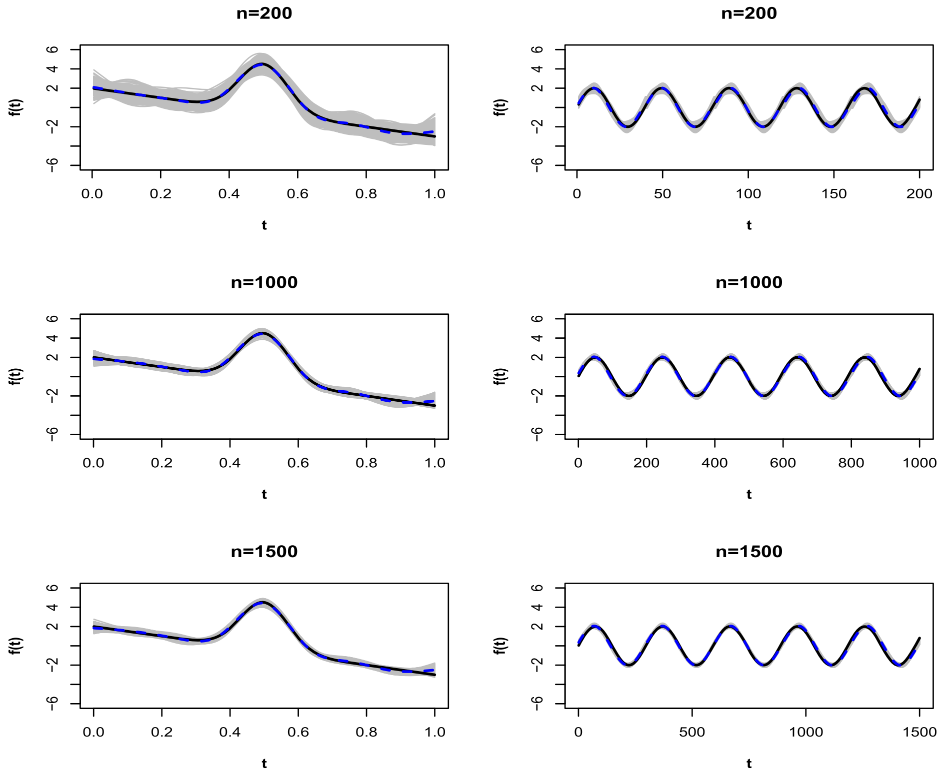


Figure 3. Graphs of the non parametric components of model (15) with 1000 replications. Adjusted curves (gray lines), mean of the estimates (blue lines), and true curves (black lines): $f_T(t)$ (first column) and $f_S(s)$ (second column).

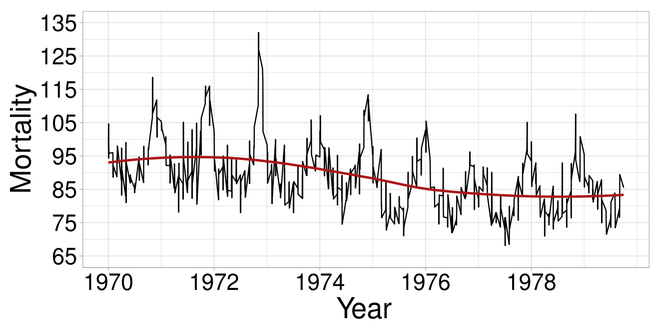


Figure 4. Time series of the weekly average cardiovascular mortality in Los Angeles from January 1970 to October 1979.

Similar to Oliveira and Paula (2021) and Wood (2017, Sec.7.7.2), we propose the following additive model to explain the average weekly cardiovascular mortality in Los Angeles:

$$\overline{\text{Mortality}}_i = f_T(\text{time}_i) + f_S(\text{week.of. year}_i) + \epsilon_i, \tag{16}$$

where $f_T(\cdot)$ and $f_S(\cdot)$ are smooth functions approximated by cubic regression and cyclic cubic regression splines, respectively, whereas $\overline{\text{Mortality}}_i$ denotes the average weekly cardiovascular mortality of the ith week, time_i is the ith week, week.of.year_i denotes the week of the year

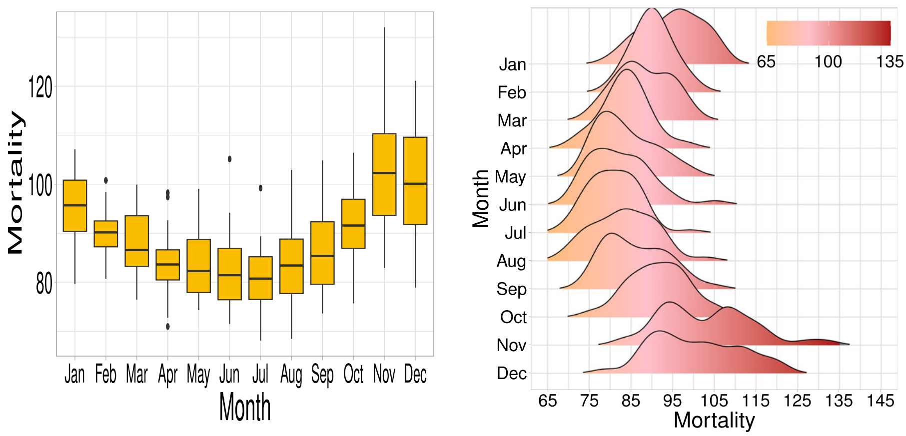


Figure 5. Boxplot and empirical distribution of the weekly average cardiovascular mortality in Los Angeles.

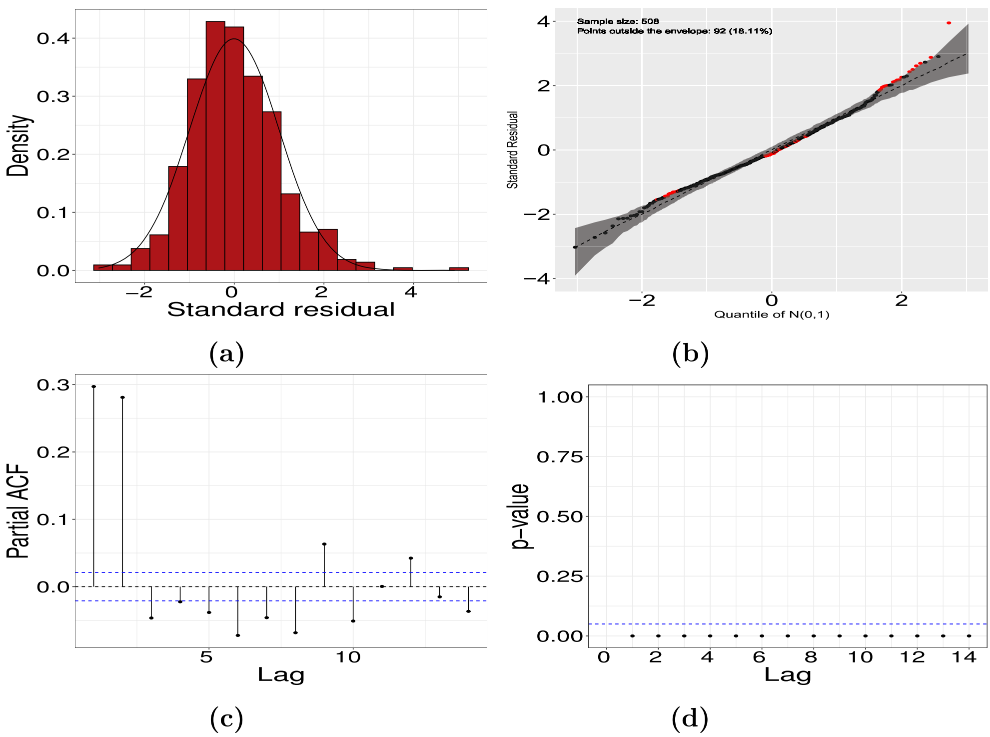


Figure 6. Residual index plots (a), normal probability plots (b), PACF (c), and Ljung-Box (d) from model (16) under independent normal error structure fitted to the weekly average cardiovascular mortality in Los Angeles.

respective to the ith week, and ϵ_i are the random errors, for $i = 1, \dots, 508$. Since one has some leap years, the default is week.of year runs from 1 to 52. We fixed the number of knots, and for this application, we selected 9 and 7 knots from time and week of the year, respectively.

Now, we assume that ϵ_i follows an AR(p) conditional asymmetric error, as defined in Section 2. Homogeneity of seasonality is also assumed over the years. Under model (16), we

Table 2. Goodness-of-fit measures of model (16) under normal, Student−t symmetric errors, and skew-normal error with AR(1), AR(2), and AR(3) structures fitted to the average weekly cardiovascular mortality in Los Angeles.

AR	Model	log-lik($\boldsymbol{\theta}$)	BIC($\boldsymbol{\theta}$)
1	N	-1576.9	3253.5
	t_{12}	-1575.1	3254.0
	SN	-1571.1	3248.2
2	N	-1556.0	3217.8
	t_{12}	-1552.4	3216.9
	SN	-1550.6	3213.3
3	N	-1555.4	3222.9
	t_{12}	-1551.9	3222.2
	SN	-1550.2	3218.6

Table 3. Parametric parameter estimates (approximate standard errors) of model (16) under normal, Student-t symmetric errors, and skew-normal error with AR(1), AR(2), and AR(3) structures fitted to the average weekly cardiovascular mortality in Los Angeles.

AR	Model	$\hat{\sigma}^2$	$\hat{\delta}$	$\hat{\psi}_1$	$\hat{\psi}_2$	$\hat{\psi}_3$
1	N	29.046 (1.823)	-	0.297(0.042)	-	-
	t_{12}	24.583(1.766)	-	0.329(0.037)	-	-
	SN	14.003(2.754)	6.436(0.692)	0.259(0.043)	-	-
2	N	26.750(1.679)	-	0.214(0.043)	0.282(0.043)	-
	t_{12}	22.289(1.611)	-	0.206(0.040)	0.276(0.044)	
	SN	13.285(2.265)	6.088(0.697)	0.184(0.043)	0.272(0.042)	-
3	N	26.691(1.675)	-	0.227(0.044)	0.292(0.043)	-0.047(0.044)
	t_{12}	22.257(1.610)	-	0.219(0.041)	0.284(0.045)	-0.044(0.044)
	SN	13.323(2.673)	6.067(0.700)	0.194(0.045)	0.280(0.042)	-0.040(0.043)

want to control the autocorrelation and seasonality properly and then detect the trend of the weekly mortality over decades.

We apply the BIC described in Section 3.1 to select the appropriate smoothing parameter values (λ_T, λ_S), with the interval $(0.1, 25)$ for λ_T and $(0.01, 10)$ for λ_S. Table 2 describes the log-likelihood and BIC values of model (16) under normal, Student−t and skew-normal AR errors, with AR(1), AR(2), and AR(3) structures, fitted to the average weekly cardiovascular mortality in Los Angeles. The AR(2) skew-normal error model presents the smallest BIC value.

The parametric parameter estimates from the fitted models are described in Table 3. The autocorrelation coefficient estimates and their approximate standard errors are very close, but the scale parameter estimates are not comparable since they are in different scales. In the sequel, we compare the fitted AR(2) error models according to the residual analysis and sensitivity studies.

Figure 7 describes the Ljung-Box tests showing that the time series is not white noise because some of the autocorrelations are not significant for all three models. Also, the autocorrelation functions (ACFs) and the partial ACFs from the three series of conditional quantile residuals confirm the adequacy of AR(2) errors in the three models.

Figure 8 presents the conditional quantile residual index plots of model (16) under normal, Student−t, and skew-normal errors with AR(2) structure fitted to the data. The index plots do not present any trend or variability over time, which indicates that the error variance is constant and that trend and seasonality were controlled. But, when considering the histograms of the normal and Student-t errors, positive asymmetry can be observed,

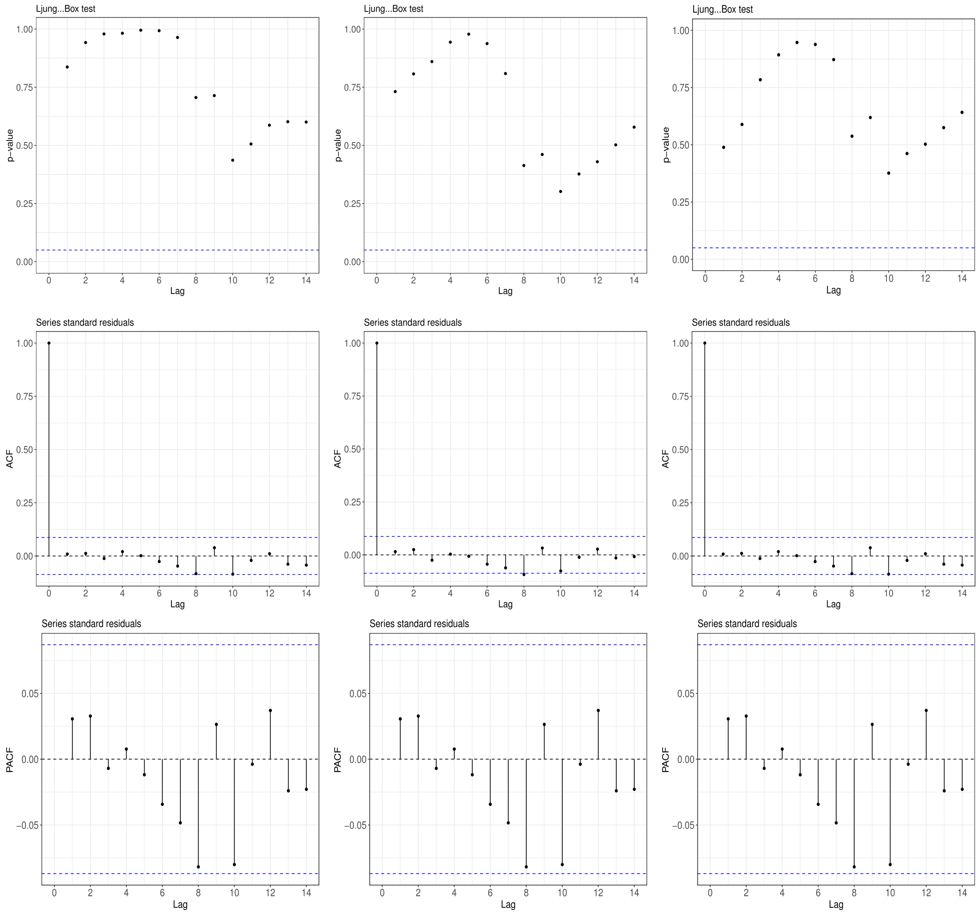


Figure 7. Ljung-Box test (top), autocorrelation functions (middle) and partial autocorrelation functions (bottom) of model (16) under normal (left), Student-*t* (middle) and Skew-Normal model (right) errors with AR(2) structure fitted to the average weekly cardiovascular mortality in Los Angeles.

which is solved by the skew-normal error. The sample skewness coefficients of the normal, Student-t, and skew-normal distributions were 0, 36, 0, 26, and 0, 01, respectively. Finally, the simulated envelopes indicate that the skew-normal model presents the best results presented (no observation outside of the confidence bands).

Finally, Figure 9 depicts the pointwise confidence bands for the trend and seasonality (annual cycle) of the average weekly cardiovascular mortality in Los Angeles from the chosen model. The left panel indicates a continuous drop in mortality, but an upward trend in the last two years. The panel on the right indicates a low mortality rate in the months from January to August and high mortality in the other months, with little variability.

Concerning the sensitivity studies, Figure 10 presents local influence diagnostic analysis using the perturbation schemes of case-weight and response variable perturbations. For the perturbation schemes, we calculated the values of $M(0)$ and plot their index graphs. The horizontal lines delimit the Lee and Xu (2004) benchmark for $M(0)$, with $c^* = 3$. For case-weight perturbation, observation #91 appears as possibly influential while #91, #110 and #363

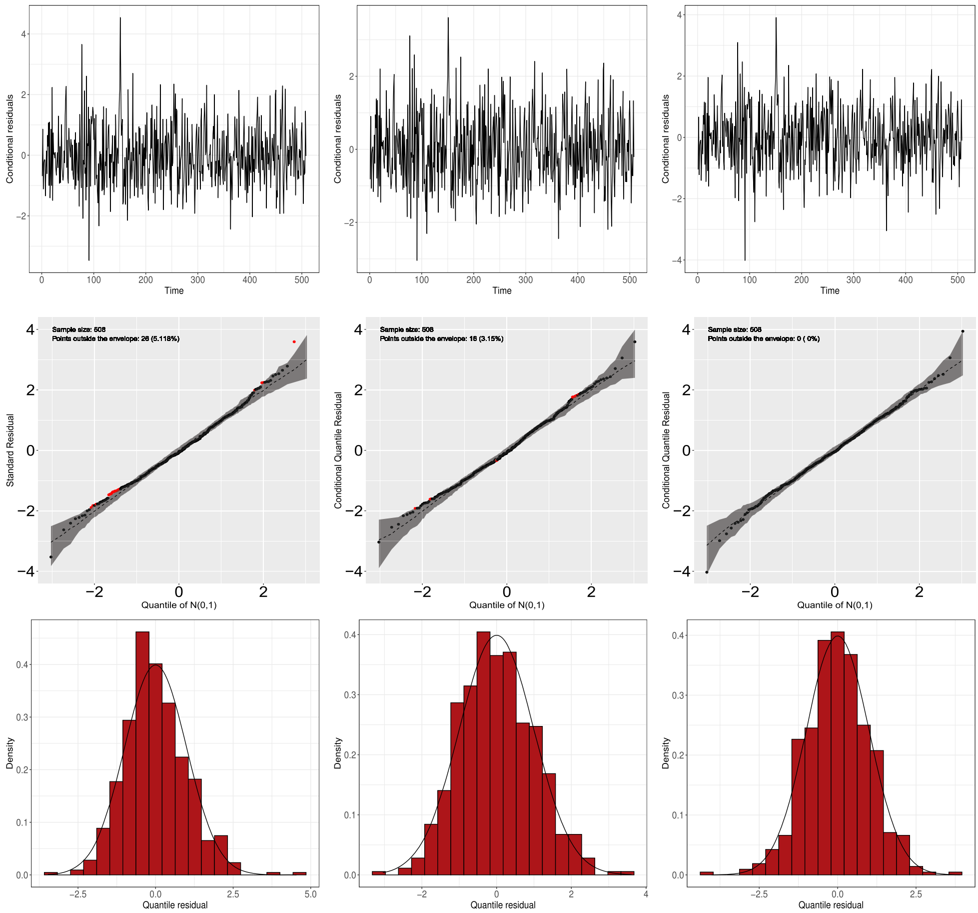
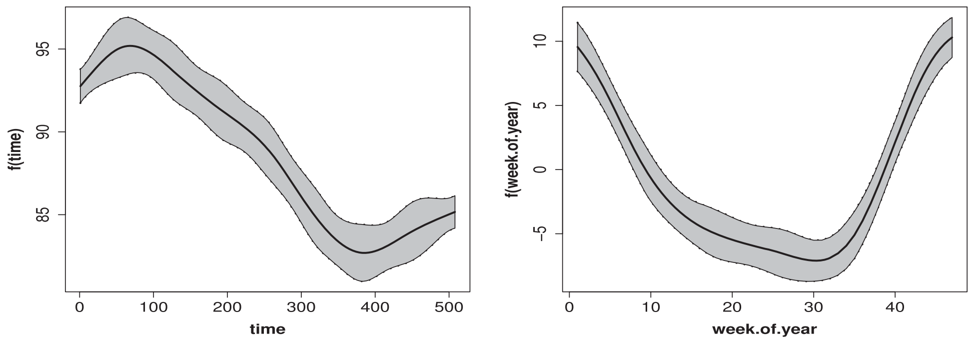


Figure 8. Residual index plots (top), normal probability plots (middle) and density of model (16) under normal (left), Student-*t* (middle) and Skew-Normal model (right) errors with AR(2) structure fitted to the weekly average cardiovascular mortality in Los Angeles.

are influences for response variable perturbation. These points are the observations with the largest negative residuals.

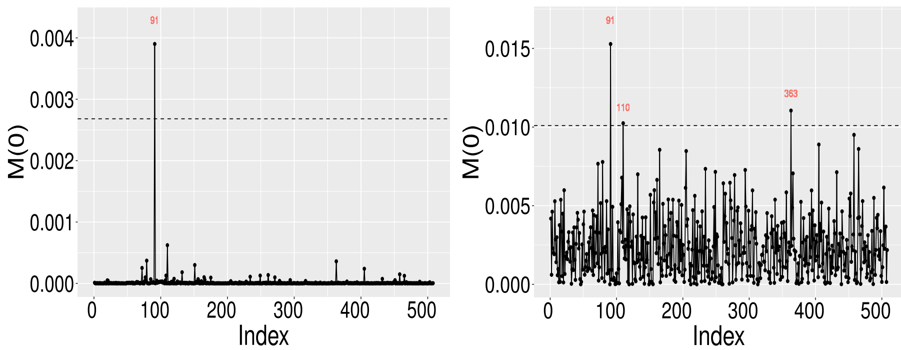
7. Final considerations

In this work, we derived parameter estimation and local influence analysis for additive models with p-order autoregressive errors under a skew-normal distribution for modeling trend and seasonality in time series. The EM algorithm for estimating the parameters of the model has analytic expressions for the M and E-steps. The effective degrees of freedom are calculated based on the linear predictor corresponding to non parametric fitting which is interpreted as linear smoothing. We developed local influence analysis based on Zhu and Lee's approach (Zhu and Lee 2001) under case-weight and response variable perturbations. We performed a simulation study considering the Doppler effect and the sine function for the non parametric components. The results suggest that the EM estimators, for a large sample, seem unbiased and



(a) (b)

Figure 9. Average weekly cardiovascular mortality in Los Angeles under ASN-AR(2) model: Partial residuals and approximate 95% pointwise confidence band for the non parametric functions for time (a); and day of the year (b).



(a) (b)

Figure 10. Average weekly cardiovascular mortality in Los Angeles under the ASN-AR(2) model: Diagnostics of case weight perturbation (a) and response variable perturbation (b).

consistent. The model is applied to the average weekly cardiovascular mortality in Los Angeles from January 1970 to October 1979, showing the usefulness of ASN-AR(p) to fit datasets with non parametric components in which the responses are asymmetric with dependence on time. The R codes (R Core Team 2023) developed in this work are available upon request. Future work would be to propose an additive model with p-order autoregressive errors using an appropriate skew level of a heavy-tailed family of distributions, such as the skew-t distribution.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The first author was supported by FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) Grant CEX-APQ-00409-22. The second author was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) Grant 305189/2018-2.

References

Azzalini, A. 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12:171–8.

Bayes, C. L., and M. D. Branco. 2007. Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics* 21:141–63.

Box, G., G. Jenkins, and G. Reinsel, 2008. *Time series analysis: Forecasting and control*, 4 ed. Oxford: Wiley.

Cao, C.-Z., J.-G. Lin, and L.-X. Zhu. 2010. Heteroscedasticity and/or autocorrelation diagnostics in nonlinear models with AR(1) and symmetrical errors. *Statistical Papers* 51 (4):813–36. doi: 10.1007/s00362-008-0171-y.

Cook, R. D. 1986. Assessment of local influence. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 48 (2):133–55. doi: 10.1111/j.2517-6161.1986.tb01398.x.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 39 (1):1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x.

Dunn, P. K., and G. K. Smyth. 1996. Randomized quantile residuals. *Journal of Computational and Graphical Statistics* 5 (3):236–44. doi: 10.1080/10618600.1996.10474708.

Eilers, P. H. C., and B. D. Marx. 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* 11:89–121.

Engle, R. F., C. W. J. Granger, J. Rice, and A. Weiss. 1986. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association* 81 (394):310–20. doi: 10.1080/01621459.1986.10478274.

Faraway, J. J. 2016. *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*, 2 ed. Boca Raton, FL: CRC Press.

Ferreira, C. S., F. Vilca, and H. Bolfarine. 2018. Diagnostics analysis for skew-normal linear regression models: Applications to a quality of life dataset. *Brazilian Journal of Probability and Statistics* 32:525–44.

Ferreira, G., L. M. Castro, V. H. Lachos, and R. Dias. 2013. Bayesian modeling of autoregressive partial linear models with scale mixture of normal errors. *Journal of Applied Statistics* 40 (8):1796–816. doi: 10.1080/02664763.2013.796349.

Ferreira, J. T. A. S, and M. F. J. Steel. 2006. A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association* 101 (474):823–9. doi: 10.1198/016214505000001212.

Genton, M. G. 2004. *Skew-elliptical distributions and their applications: A journey beyond normality*. London: Chapman & Hall.

Green, P. J. 1987. Penalized likelihood for general semi-parametric regression models. *International Statistical Review / Revue Internationale De Statistique* 55 (3):245 doi: 10.2307/1403404.

Green, P. J. 1990. On use of the em algorithm for penalized likelihood estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 52 (3):443–52. doi: 10.1111/j.2517-6161.1990.tb01798.x.

Harezlak, J., D. Ruppert, and M. P. Wand. 2018. *Semiparametric regression with R*. New York: Springer.

Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller. 2010. Portfolio selection with higher moments. *Quantitative Finance* 10 (5):469–85. doi: 10.1080/14697681003756877.

Hastie, T., and R. Tibshirani, 1990. *Generalized additive models*. London: Chapman and Hall.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions*, Vol. 1. New York: John Wiley.

Lee, S.-Y., and L. Xu. 2004. Influence analyses of nonlinear mixed-effects models. *Computational Statistics & Data Analysis* 45 (2):321–41. doi: 10.1016/S0167-9473(02)00303-1.

Lin, W., and K. B. Kulasekera. 2007. Identifiability of single-index models and additive-index models. *Biometrika* 94 (2):496–501. doi: 10.1093/biomet/asm029.

Liu, S. 2004. On diagnostics in conditionally heteroskedastic time series models under elliptical distributions. *Journal of Applied Probability* 41 (A):393–405. doi: 10.1239/jap/1082552214.

Nadarajah, S., and S. Kotz. 2003. Skewed distributions generated by the normal kernel. *Statistics & Probability Letters* 65 (3):269–77. doi: 10.1016/j.spl.2003.07.013.
Oliveira, R. A., and G. A. Paula. 2021. Additive models with autoregressive symmetric errors based on penalized regression splines. *Computational Statistics* 36 (4):2435–66. doi: 10.1007/s00180-021-01106-2.
Paula, G. A., M. Medeiros, and F. E. Vilca-Labra. 2009. Influence diagnostics for linear models with first-order autoregressive elliptical errors. *Statistics & Probability Letters* 79 (3):339–46. doi: 10.1016/j.spl.2008.08.017.
R Core Team. 2023. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.
Racine, J. S., L. Su, and A. Ullah, 2014. *The Oxford handbook of applied nonparametric and semiparametric econometrics and statistics*. New York: Oxford University Press.
Relvas, C. E. M., and G. A. Paula. 2016. Partially linear models with first-order autoregressive symmetric errors. *Statistical Papers* 57 (3):795–825. doi: 10.1007/s00362-015-0680-4.
Sahu, S. K., D. K. Dey, and M. D. Branco. 2003. A new class of multivariate skew distributions with applications to bayesian regression models. *Canadian Journal of Statistics* 31 (2):129–50. doi: 10.2307/3316064.
Segal, M. R., P. Bacchetti, and N. P. Jewell. 1994. Variances for maximum penalized likelihood estimates obtained Via the Em algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 56 (2):345–52. doi: 10.1111/j.2517-6161.1994.tb01983.x.
Shumway, R. H. 1988. *Applied statistical time series analysis*. Englewood Cliffs, NJ: Prentice-Hall.
Stringer, A. 2024. Identifiability constraints in generalized additive models. *Canadian Journal of Statistics* 52 (2):461–76. doi: 10.1002/cjs.11786.
Wood, S. 2017. *Generalized additive models, an Introduction with R, 2nd ed.* Boca Raton: Chapman and Hall.
Yu, Y., and D. Ruppert. 2002. Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association* 97 (460):1042–54. doi: 10.1198/016214502388618861.
Yuan, M. 2011. On the identifiability of additive index models. *Statistica Sinica* 21:1901–11.
Zheng, S., and D. Li. 2018. Semiparametric time series regression modeling with a diverging number of parameters. *Statistica Neerlandica* 72 (2):90–108. doi: 10.1111/stan.12121.
Zhu, H.-T., and S.-Y. Lee. 2001. Local influence for incomplete data models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 63 (1):111–26. doi: 10.1111/1467-9868.00279.

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 usual likelihood Segal, Bacchetti, and Jewell (1994). Given the observed log-likelihood function $\ell(\boldsymbol{\theta})$ in (5), the corresponding penalized log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\gamma}_T^\top, \boldsymbol{\gamma}_S^\top, \sigma^2, \delta, \boldsymbol{\psi})^\top$ has the form:

$$\ell_p(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \frac{\alpha_T}{2}\boldsymbol{\gamma}_T^\top \mathbf{K}_T \boldsymbol{\gamma}_T - \frac{\alpha_S}{2}\boldsymbol{\gamma}_S^\top \mathbf{K}_S \boldsymbol{\gamma}_S. \tag{A.1}$$

The observed information matrix for $\boldsymbol{\theta}$ can 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\frac{\partial B_i}{\partial \boldsymbol{\gamma}_T} = -\frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}(\mathbf{n}_{T_i} - \sum_{j=1}^{p} \psi_j \mathbf{n}_{T_{i-j}}),\ \frac{\partial B_i}{\partial \boldsymbol{\gamma}_S} = -\frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}(\mathbf{n}_{S_i} - \sum_{j=1}^{p} \psi_j \mathbf{n}_{S_{i-j}}),\ \frac{\partial B_i}{\partial \psi_j} =$$
$$-\frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}(y_{i-1} - \mu_{i-j}), \frac{\partial B_i}{\partial \sigma^2} = -\frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3(\sigma^2+\delta^2)^{3/2}}(y_i - \xi_i + b\delta), \frac{\partial B_i}{\partial \delta} = \frac{\sigma^2(y_i-\xi_i+b\delta)+b\delta(\sigma^2+\delta^2)}{\sigma(\sigma^2+\delta^2)^{3/2}},$$
$$\frac{\partial^2 B_i}{\partial \boldsymbol{\gamma}_T \partial \boldsymbol{\gamma}_T^\top} = \frac{\partial^2 B_i}{\partial \boldsymbol{\gamma}_T \partial \boldsymbol{\gamma}_S^\top} = \frac{\partial^2 B_i}{\partial \boldsymbol{\gamma}_S \partial \boldsymbol{\gamma}_S^\top} = \mathbf{0},\ \frac{\partial^2 B_i}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^\top} = \mathbf{0},$$
$$\frac{\partial^2 B_i}{\partial \psi_j \partial \boldsymbol{\gamma}_T} = \frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}\mathbf{n}_{T_{i-j}}, \frac{\partial^2 B_i}{\partial \psi_j \partial \boldsymbol{\gamma}_S} = \frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}\mathbf{n}_{S_{i-j}},$$
$$\frac{\partial^2 B_i}{\partial \sigma^2 \partial \boldsymbol{\gamma}_T} = \frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3(\sigma^2+\delta^2)^{3/2}}(\mathbf{n}_{T_i} - \sum_{j=1}^{p} \psi_j \mathbf{n}_{T_{i-j}}),\ \frac{\partial^2 B_i}{\partial \sigma^2 \partial \boldsymbol{\gamma}_S} = \frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3(\sigma^2+\delta^2)^{3/2}}(\mathbf{n}_{S_i} - \sum_{j=1}^{p} \psi_j \mathbf{n}_{S_{i-j}}),$$
$$\frac{\partial^2 B_i}{\partial \sigma^2 \partial \psi_j} = \frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3(\sigma^2+\delta^2)^{3/2}}(y_{i-j} - \mu_{i-j}),\quad \frac{\partial^2 B_i}{\partial \delta \partial \boldsymbol{\gamma}_T} = -\frac{\sigma}{(\sigma^2+\delta^2)^{3/2}}(\mathbf{n}_{T_i} - \sum_{j=1}^{p} \psi_j \mathbf{n}_{T_{i-j}}),\quad \frac{\partial^2 B_i}{\partial \delta \partial \boldsymbol{\gamma}_S} =$$
$$-\frac{\sigma}{(\sigma^2+\delta^2)^{3/2}}(\mathbf{n}_{S_i} - \sum_{j=1}^{p} \psi_j \mathbf{n}_{S_{i-j}}), \frac{\partial^2 B_i}{\partial \delta \partial \psi_j} = -\frac{\sigma}{(\sigma^2+\delta^2)^{3/2}}(y_{i-j} - \mu_{i-j}),$$
$$\frac{\partial^2 B_i}{\partial \sigma^4} = -\frac{\delta(y_i - \xi_i + b\delta)}{4\sigma^5(\sigma^2+\delta^2)^{5/2}}[4\sigma^2(\sigma^2+\delta^2) - 3(2\sigma^2+\delta^2)^2],$$
$$\frac{\partial^2 B_i}{\partial \sigma^2 \partial \delta} = \frac{\sigma(\sigma^2+\delta^2)(y_i - \xi_i + 2b\delta) - \frac{1}{2}\left[\sigma^2(y_i-\xi_i+b\delta)+b\delta(\sigma^2+\delta^2)\right]\left(\frac{\sigma^2+\delta^2}{\sigma} + 3\sigma\right)}{\sigma^2(\sigma^2+\delta^2)^{5/2}},$$
$$\frac{\partial^2 B_i}{\partial \delta^2} = \frac{b(\sigma^2+\delta^2)(2\sigma^2+3\delta^2) - 3\delta\left[\sigma^2(y_i-\xi_i+b\delta)+b\delta(\sigma^2+\delta^2)\right]}{\sigma(\sigma^2+\delta^2)^{5/2}},$$

where $\mathbf{x}_0 = ... = \mathbf{x}_{-(p-1)} = \mathbf{0}$, $\mathbf{n}_0 = ... = \mathbf{n}_{-(p-1)} = \mathbf{0}$, $y_0 = ..., y_{-(p-1)} = 0$ and $\mu_0 = ... = \mu_{-(p-1)} = 0$.