



UNIVERSITY OF LEEDS

This is a repository copy of *A new BISARMA time series model for forecasting mortality using weather and particulate matter data*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/162235/>

Version: Accepted Version

Article:

Leiva, V, Saulo, H, Souza, R et al. (2 more authors) (2021) A new BISARMA time series model for forecasting mortality using weather and particulate matter data. *Journal of Forecasting*, 40 (2). pp. 346-364. ISSN 0277-6693

<https://doi.org/10.1002/for.2718>

© 2020 John Wiley & Sons, Ltd. This is the peer reviewed version of the following article: Leiva, V, Saulo, H, Souza, R et al. (2 more authors) (2021) A new BISARMA time series model for forecasting mortality using weather and particulate matter data. *Journal of Forecasting*, 40 (2). pp. 346-364. ISSN 0277-6693, which has been published in final form at <http://doi.org/10.1002/for.2718>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

A new BISARMA time series model for forecasting mortality using weather and particulate matter data

Víctor Leiva^{1*}, Helton Saulo², Rubens Souza², Robert G. Aykroyd³, Roberto Vila²

¹School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Chile

²Department of Statistics, Universidade de Brasília, Brazil

³Department of Statistics, University of Leeds, United Kingdom

Abstract

The Birnbaum-Saunders (BS) distribution is a model that frequently appears in the statistical literature and has proved to be very versatile and efficient across a wide range of applications. However, despite the growing interest in the study of this distribution and the development of many review articles, few papers have considered data with a dependency structure. To fill this gap, we introduce a new class of time-series models based on the BS distribution, which allows modelling of non-negative and asymmetric data that have an autoregressive structure. We call these Birnbaum-Saunders autoregressive moving averaging (BISARMA) models. Also included is a thorough study of theoretical properties of the proposed methodology and of practical issues, such as maximum likelihood parameter estimation, diagnostic analysis and prediction. The performance of the proposed methodology is evaluated using a Monte Carlo simulation. Finally, **an analysis** of real-world mortality data is performed using the methodology to show its potential for applications. The numerical results show the excellent performance of the BISARMA model, indicating that the BS distribution is a good modeling choice when dealing with time series data which are both non-negative and asymmetric, and hence it can be a valuable addition to the tool-kit of applied statisticians and data scientists.

Keywords ARMA models; Birnbaum-Saunders distribution; data dependent over time; maximum likelihood and Monte Carlo methods; model selection; residuals; R software.

1 Introduction

The Birnbaum-Saunders (BS) distribution frequently arises in the applied statistical literature. In the last decades, it has been widely studied and has been shown to be versatile and efficient in several fields of science, due to theoretical justification, its good properties and its close relationship with the normal distribution. The BS distribution is defined on the positive real numbers, is unimodal with positive skewness and has two parameters that control its shape and scale. It is often considered as a life distribution due to its origins describing fatigue of materials **subject to stress**. Hence, it has a prominent role in the areas of reliability and survival analysis, being a good alternative to symmetric distributions. For more details, see Birnbaum and Saunders (1969), Johnson et al. (1995, pp. 651-663) and Leiva (2016).

*Corresponding author: Victor Leiva. Email: victorleivasanchez@gmail.com. URL: <http://www.victorleiva.cl>

Although the BS distribution has its origins in physics and engineering, it has also received interest in several other areas. In particular, it is attracting interest in the earth and medical sciences, following its reformulation for environmental and medical processes formalized by Leiva et al. (2015) and Leão et al. (2018b), respectively. For examples of environmental and medical applications of BS models conducted by international transdisciplinary groups the reader is referred to Ferreira et al. (2012), Marchant et al. (2013, 2016a,b, 2018, 2019), Saulo et al. (2013), Leiva et al. (2016), Garcia-Papani et al. (2017, 2018a,b), Leão et al. (2017, 2018a,b), Desousa et al. (2018), Lillo et al. (2018), Huerta et al. (2019) and Martinez et al. (2019).

The study of the BS distribution has received growing interest and a considerable body of work is available, with recent publications by Leiva (2016), Aykroyd et al. (2018), and Balakrishnan and Kundu (2019) providing a thorough summary. To date, however, there has been little development of methods for the analysis of data with temporal dependence structure, and particularly for time series, based on the BS distribution. Some efforts along these line are attributed to Bhatti (2010), Leiva et al. (2014) and Saulo et al. (2019) who, motivated by the work of Engle and Russell (1998), developed autoregressive conditional duration models based on the BS distribution. Other recent studies related to BS autoregressive models are attributed to Fonseca and Cribari-Neto (2018) and Rahul et al. (2018).

Following on from the above, the main objective of this work is to propose a new class of models based on the BS distribution for time series data, named the Birnbaum-Saunders autoregressive moving averaging (BISARMA) model. Our proposal can be thought of as analogous to ARMAX models; see Shumway and Stoffer (2017) and it is an extension of the work of Benjamin et al. (2003), Rocha and Cribari-Neto (2009), and Maior and Cysneiros (2016). The authors extended the autoregressive moving average (ARMA) time series model to different classes of distributions, such as the generalized ARMA (GARMA) model (see Ben Amor et al., 2018), which **includes** the exponential family, beta-ARMA for beta distributions and SYMARMA for symmetric distributions. We apply the BISARMA model for forecasting mortality using weather and particulate matter data; see Xu (2020).

After this introduction, the paper is organized as follows. Section 2 defines the BS distribution, its logarithmic version (log-BS) and some of their properties. In Section 3, we formulate the BISARMA model and provide estimation, prediction and residual analysis based on maximum likelihood (ML). In Section 4, a Monte Carlo simulation study is reported to evaluate the performance of the proposed methodology. The section also describes analytics of real mortality time series data, including a diagnostic analysis, to show the potential of the proposed methods. Section 5 gives concluding remarks. Appendixes with detailed mathematical proofs and expressions are also presented.

2 Basic definitions

2.1 The Birnbaum-Saunders distribution

If a random variable T follows a BS distribution, usually denoted by $T \sim \text{BS}(\alpha, \beta)$, then its distribution can be represented by the cumulative distribution function (CDF) given by

$$F_T(t; \alpha, \beta) = \Phi \left[\frac{1}{\alpha} \left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \right], \quad t > 0, \alpha > 0, \beta > 0, \quad (1)$$

where Φ is the standard normal distribution function, α is a shape parameter, **and β is a scale parameter that is also the median of the distribution.** The probability density function (PDF) of T is obtained by

differentiating (1) with respect to t , that is, $f_T(t; \alpha, \beta) = F_T'(t; \alpha, \beta)$, so that

$$f_T(t; \alpha, \beta) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2\alpha^2} \left(\frac{t}{\beta} + \frac{\beta}{t} - 2 \right) \right] \frac{t^{-3/2}}{2\alpha\sqrt{\beta}}(t + \beta), \quad t > 0, \alpha > 0, \beta > 0.$$

Figure 1 shows examples of the BS PDF. Note that, in Figure 1(a), as the value of α increases, the degree of asymmetry increases, causing the variance to increase and also the flattening of the density function. When α approaches zero, the curve approaches the symmetric case, centred around the fixed value β (the median of the distribution), with $\beta = 1$ here, and its variability decreases. In Figure 1(b), setting the value $\alpha = 0.1$ and varying β , the shape of the PDF does not change and only its location and scale changes; that is, the PDF moves and its spread changes in the same sense as the change of β . This shows the fact that β is both a location parameter (defining the median of the distribution) and a scale parameter. **The BS PDF is always unimodal, see, for example, Proposition 2.7 in Vila et al.**

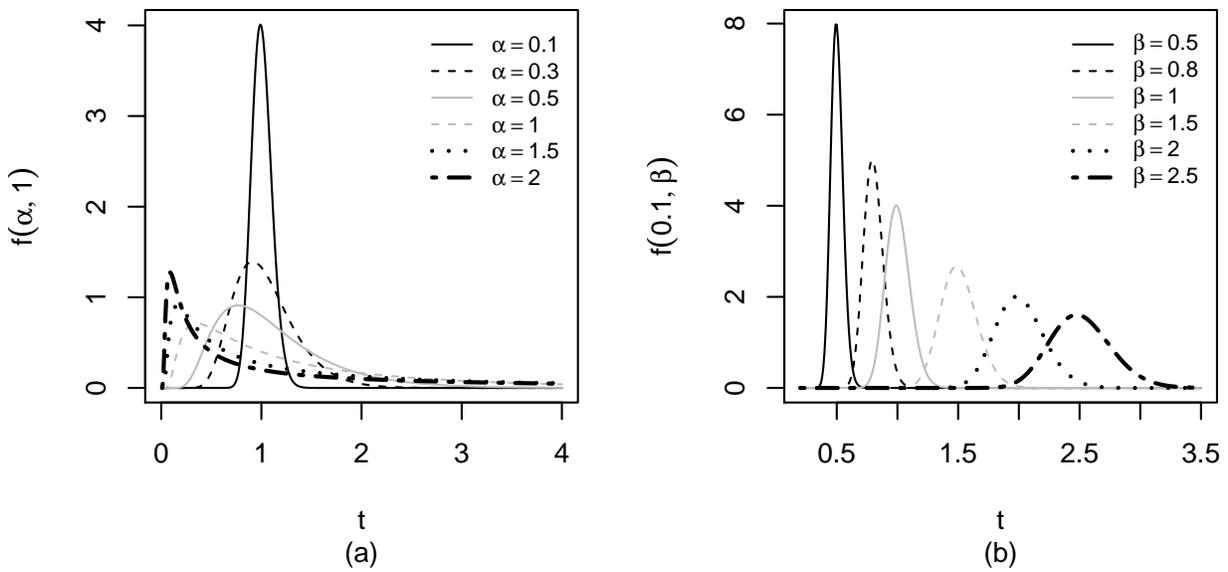


Figure 1: PDF of the distribution: (a) $BS(\alpha, 1)$ for selected values of the shape parameter α , and (b) $BS(0.1, \beta)$ for selected values of the scale parameter β .

(2020).

An important and well-known property in the construction of this distribution is that a random variable $T \sim BS(\alpha, \beta)$ can be generated from a random variable Z , which has the standard normal distribution, through the relationship

$$T = \beta \left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2} \right)^2 + 1} \right]^2,$$

where

$$Z = \frac{1}{\alpha} \left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right) \sim N(0, 1).$$

This relationship is extremely useful and gives a method to obtain pseudo-random numbers from the BS distribution; [see details in Leiva \(2016\)](#).

Some properties of the BS distribution are now presented. If $T \sim \text{BS}(\alpha, \beta)$, then:

- (i) for any real constant $k > 0$, we have $kT \sim \text{BS}(\alpha, k\beta)$, which means that the BS distribution is closed under proportionality;
- (ii) $1/T \sim \text{BS}(\alpha, 1/\beta)$, implying that the BS distribution is also closed under reciprocity;
- (iii) the parameter β is a scale parameter, hence $T/\beta \sim \text{BS}(\alpha, 1)$. Furthermore, β is the median of the distribution, which can be obtained directly by making $q = 0.5$ in the quantile function which is given by

$$t(q; \alpha, \beta) = F_T^{-1}(q; \alpha, \beta) = \beta \left[\frac{\alpha z(q)}{2} + \sqrt{\left(\frac{\alpha z(q)}{2} \right)^2 + 1} \right]^2, \quad 0 < q < 1,$$

where z is the standard normal quantile function and F_T^{-1} is the inverse of F_T defined in (1);

- (iv) the parameter α is a shape parameter, such that when $\alpha \rightarrow 0$, the BS distribution tends to a normal distribution $\text{N}(\beta, \tau)$, [where \$\tau \rightarrow 0\$](#) ;

(v) expressions for the BS mean and variance are hence given by

$$\text{E}[T] = \beta \left(1 + \frac{1}{2}\alpha^2 \right) \quad \text{and} \quad \text{Var}[T] = (\alpha\beta)^2 \left(1 + \frac{5}{4}\alpha^2 \right)$$

and for the coefficients of skewness and kurtosis by

$$\text{CS}[T] = 16\alpha^2 \frac{(11\alpha^2 + 6)}{(5\alpha^2 + 4)^3} \quad \text{and} \quad \text{CK}[T] = 3 + 6\alpha^2 \frac{(93\alpha^2 + 41)}{(5\alpha^2 + 4)^2}.$$

Consequently, according to the property of reciprocity, we have that

$$\text{E}[T^{-1}] = \beta^{-1} \left(1 + \frac{1}{2}\alpha^2 \right) \quad \text{and} \quad \text{Var}[T^{-1}] = (\alpha\beta^{-1})^2 \left(1 + \frac{5}{4}\alpha^2 \right).$$

2.2 The log-Birnbaum-Saunders distribution

The log-BS distribution, denoted as $\text{log-BS}(\alpha, \mu)$, is derived as the logarithm of a $\text{BS}(\alpha, \beta)$ random variable. Rieck and Nedelman (1991) proved that if $T \sim \text{BS}(\alpha, \beta)$, then $Y = \log(T) \sim \text{log-BS}(\alpha, \mu)$ with shape and location parameters given by $\alpha > 0$ and $\mu = \log(\beta) \in \mathbb{R}$, respectively.

In an alternative formulation, the random variable Y follows a log-BS distribution if and only if

$$Z = \frac{2}{\alpha} \sinh \left(\frac{Y - \mu}{2} \right) \sim \text{N}(0, 1).$$

Then, the CDF of Y is given by

$$F_Y(y; \alpha, \mu) = \Phi \left[\frac{2}{\alpha} \sinh \left(\frac{y - \mu}{2} \right) \right], \quad y \in \mathbb{R}, \mu \in \mathbb{R}, \alpha > 0, \quad (2)$$

and differentiating (2) with respect to y , we obtain the corresponding PDF given by

$$f_Y(y; \alpha, \mu) = \frac{1}{\alpha\sqrt{2\pi}} \exp \left[-\frac{2}{\alpha^2} \sinh^2 \left(\frac{y - \mu}{2} \right) \right] \cosh \left(\frac{y - \mu}{2} \right), \quad y \in \mathbb{R}, \mu \in \mathbb{R}, \alpha > 0.$$

The mean and variance of Y can be obtained using the moment-generating function defined as

$$M_Y(s) = \exp(\mu s) \left[\frac{K_{[\sigma s+1]/2}(\delta^{-2}) + K_{[\sigma s-1]/2}(\delta^{-2})}{2K_{1/2}(\delta^{-2})} \right],$$

where K_λ is a modified Bessel function of the third kind given by

$$K_\lambda(w) = \frac{w^\lambda}{2^{\lambda+1}} \int_0^\infty u^{-\lambda-1} \exp[-u - (w^2/4u)] \, du.$$

Some properties of the log-BS distribution are now presented. If $Y \sim \text{log-BS}(\alpha, \mu)$, then:

- (i) $T = \exp(Y) \sim \text{BS}(\alpha, \beta)$;
- (ii) $E[Y] = \mu = \log(\beta)$;
- (iii) there is no closed form for the variance of Y , but, based on an asymptotic approximation for the moment-generating function of the log-BS distribution, it follows that if $\alpha \rightarrow 0$, then $\text{Var}[T] = \alpha^2 - \alpha^4/4$, whereas if $\alpha \rightarrow \infty$, then $\text{Var}[T] = 4(\log^2 \sqrt{2}\alpha) + 2 - 2\log(\sqrt{2}\alpha)$;
- (iv) if $X = Y \pm k$, then $X \sim \text{log-BS}(\alpha, \mu \pm k)$; and
- (v) the log-BS distribution is symmetric around μ , unimodal for $0 < \alpha \leq 2$ and bimodal for $\alpha > 2$.

Another important property of the BS distribution is that its logarithmic version has flexible bimodality. Note that if $T \sim \text{BS}(\alpha, \beta)$, then $Y = \log(T) \sim \text{log-BS}(\alpha, \mu)$, where $\mu = \log(\beta)$. For more details, see Desousa et al. (2018), Leão et al. (2017, 2018a,b) and references therein.

In Figure 2(a) some of the properties of the log-BS distribution described above can be seen. **Notice that α modifies the shape of the distribution and when α increases, the kurtosis of the distribution also increases, so that the distribution becomes flatter.** In particular, for $\alpha \leq 2$, the distribution is unimodal and exhibits smaller kurtosis than the normal distribution. However, if $\alpha > 2$, then the distribution begins to display bimodality, with widely separated modes and its kurtosis is greater than that of the normal distribution. Figure 2(b) shows how μ modifies the location.

3 BISARMA model

3.1 Formulation

Let Y_1, \dots, Y_n be random variables defined in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for every $t = 1, \dots, n$, and $\mathcal{F}_t = \sigma(Y_1, \dots, Y_t)$ be a σ -algebra generated by Y_1, \dots, Y_t . In addition, we define $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Suppose that for each $t = 1, \dots, n$, the conditional distribution of Y_t , given the past data $\mathcal{F}_{t-1} = \{Y_{t-1}, \dots, Y_1, \mu_{t-1}, \dots, \mu_1\}$ follows a log-BS distribution, denoted by $Y_t | \mathcal{F}_{t-1} \sim \text{log-BS}(\alpha, \mu_t)$, with conditional PDF given by

$$f(y_t; \alpha, \mu_t | \mathcal{F}_{t-1}) = \frac{1}{\alpha\sqrt{2\pi}} \exp \left[-\frac{2}{\alpha^2} \sinh^2 \left(\frac{y_t - \mu_t}{2} \right) \right] \cosh \left(\frac{y_t - \mu_t}{2} \right), \quad y_t \in \mathbb{R}, \mu_t \in \mathbb{R}, \alpha > 0,$$

where α and $\mu_t = E[Y_t | \mathcal{F}_{t-1}]$ are, respectively, the shape parameter and the conditional mean of the

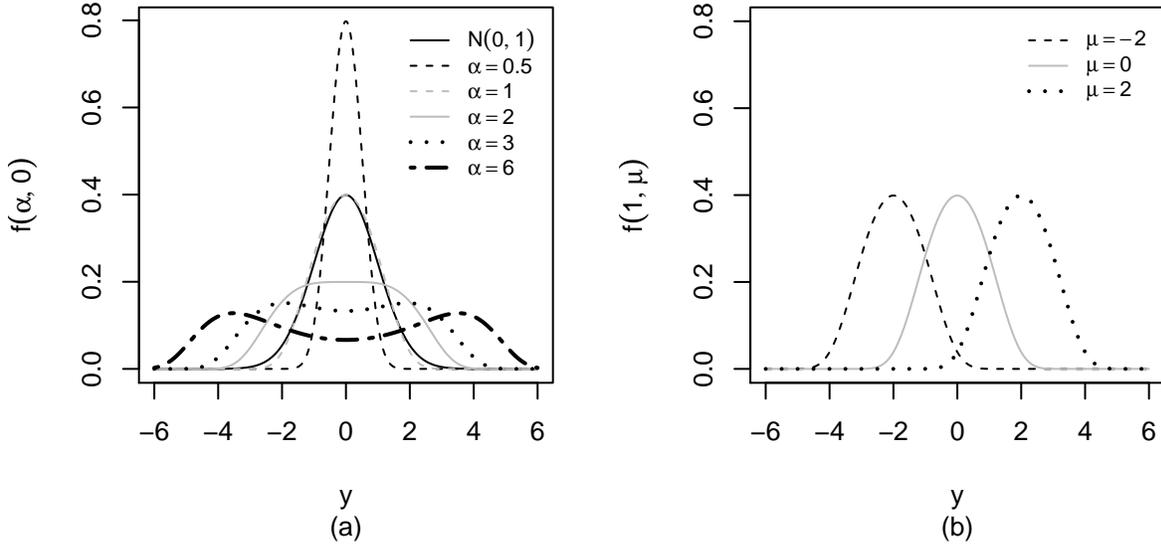


Figure 2: PDF of the log-BS distribution for (a) $\text{log-BS}(\alpha, 0)$ for selected values of the shape parameter α , and (b) $\text{log-BS}(1, \mu)$ for selected values of the location μ .

random variable Y_t .

Next, a class of BS log-linear regression models is defined by

$$Y_t = \mu_t + \varepsilon_t, \quad t = 1, \dots, n, \quad (3)$$

where $\mu_t = \mathbf{x}_t^\top \boldsymbol{\beta}$, with $\mathbf{x}_t^\top = (x_{t1}, \dots, x_{tk})$ being a vector containing the values of k explanatory variables, for $k < n$, where n is the sample size, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^\top$ is a vector of unknown parameters to be estimated and ε_t is the model error term. Note that these errors are not correlated and that $\varepsilon_t \sim \text{log-BS}(\alpha, 0)$. In the **BISARMA**(p, q) model, however, the component μ_t included in (3) contains an extra dynamic additive term, τ_t , with ARMA structure, such that now

$$\mu_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \tau_t. \quad (4)$$

The component τ_t included in (4) is defined as follows. Consider $\psi_t = Y_t - \mathbf{x}_t^\top \boldsymbol{\beta} = \tau_t + \varepsilon_t$ described by an ARMA(p, q) model expressed as

$$\psi_t = \eta + \sum_{i=1}^p \phi_i \psi_{t-i} + \sum_{j=1}^q \theta_j u_{t-j} + u_t, \quad (5)$$

where $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top \in \mathbb{R}^q$, and $p, q \in \mathbb{N}$ are the autoregressive and moving-average parameters and their respective orders, with $\eta \in \mathbb{R}$ being a constant. The terms u_t are uncorrelated random errors (white noise), which can be, for example, measurement errors on the original scale (that is, $u_t = y_t - \mu_t = \varepsilon_t$). Assuming that $E[u_t | \mathcal{F}_{t-1}] = 0$, a.s., for all t , and taking the conditional expectation, with respect to the set of past data \mathcal{F}_{t-1} , of both sides of expression (5), we

have

$$E[\psi_t | \mathcal{F}_{t-1}] = \eta + \sum_{i=1}^p \phi_i (Y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j u_{t-j}, \quad \text{a.s., for all } t.$$

Given that $E[Y_t | \mathcal{F}_{t-1}] = \mu_t$ a.s. and $\mu_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \tau_t$, we get

$$E[\psi_t | \mathcal{F}_{t-1}] = E[Y_t | \mathcal{F}_{t-1}] - \mathbf{x}_t^\top \boldsymbol{\beta} = \mu_t - \mathbf{x}_t^\top \boldsymbol{\beta} = \tau_t, \quad \text{a.s., for all } t,$$

and thus

$$\tau_t = \eta + \sum_{i=1}^p \phi_i (y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j u_{t-j}. \quad (6)$$

Therefore, from (3), (4) and (6), we propose the **BISARMA**(p, q) model defined by

$$Y_t = \eta + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i (y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j u_{t-j} + \varepsilon_t, \quad t = 1, \dots, n, \quad (7)$$

where $\varepsilon_t \sim \text{log-BS}(\alpha, 0)$, which implies that

$$Y_t = \log(T_t) \sim \text{log-BS} \left(\alpha, \eta + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i (y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j u_{t-j} \right)$$

and then

$$T_t \sim \text{BS} \left(\alpha, \exp \left[\eta + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i (y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j u_{t-j} \right] \right). \quad (8)$$

Theorem 1. Let $\Phi(B) = -\sum_{i=0}^p \kappa_i B^i$ with $\kappa_0 = -1$, and $\Theta(B) = \sum_{i=0}^q \xi_i B^i$ with $\xi_0 = 1$, be the autoregressive and moving averages polynomials respectively, where B^i is the lag operator, i.e., $B^i y_t = y_{t-i}$. Provided that $\Phi(B)$ is invertible, the marginal mean and marginal variance of Y_t , and the covariance and correlation of Y_t and Y_{t-k} in the **BISARMA**(p, q) model are given by

$$E[Y_t] = \eta + \mathbf{x}_t^\top \boldsymbol{\beta}, \quad \text{Var}[Y_t] = \sum_{i=0}^{\infty} \psi_i^2 g_{t-i}, \quad \text{Cov}[Y_t, Y_{t-k}] = \sum_{i=0}^{\infty} \psi_i \psi_{i-k} g_{t-i}, \quad k > 0,$$

$$\text{Corr}[Y_t, Y_{t-k}] = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i-k} g_{t-i}}{\prod_{j \in \{0, k\}} \sqrt{\sum_{i=0}^{\infty} \psi_i^2 g_{t-j-i}}},$$

respectively, where $g_t = E[\text{Var}[Y_t | \mathcal{F}_{t-1}]]$. See proof in Appendix A.

3.2 Estimation

Let Y_1, \dots, Y_n be a sample from the **BISARMA**(p, q) model defined in (7), and let $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}, \eta, \boldsymbol{\phi}, \boldsymbol{\theta})$ be the **parameter vector of interest**, where $m = \max\{p, q\}$ with $n > m$. For each $t = m + 1, \dots, n$, consider the log-likelihood function $\ell_t(\alpha, \boldsymbol{\beta}, \eta, \boldsymbol{\phi}, \boldsymbol{\theta}) = \log(f(y_t; \alpha, \mu_t | \mathcal{F}_{t-1}))$ given \mathcal{F}_{t-1} . Therefore,

$l_t = l_t(\alpha, \beta, \eta, \phi, \theta)$ is expressed as

$$l_t \propto -\log(\alpha) + \log \left[\cosh \left(\frac{y_t - \mu_t}{2} \right) \right] - \frac{2}{\alpha^2} \sinh^2 \left(\frac{y_t - \mu_t}{2} \right),$$

which is written as proportional to remove the components that are independent of γ and which will be eliminated by mathematical differentiation. Thus, the log-likelihood function for the parameter vector γ of the **BISARMA**(p, q) model is given by

$$\ell = \sum_{t=m+1}^n l_t(\alpha, \beta, \eta, \phi, \theta). \quad (9)$$

In order to perform the estimation of parameters in the **BISARMA**(p, q) model, we calculate the score vector, which is defined as $U(\gamma)$, formed by the first order partial derivatives, with respect to each of the parameters, of the log-likelihood function. Hence, from (9), we have

$$\frac{\partial \ell_t}{\partial \alpha} = -\frac{1}{\alpha} + \frac{4}{\alpha^3} \left[\sinh \left(\frac{y_t - \mu_t}{2} \right) \right]^2$$

and consequently

$$\frac{\partial \ell}{\partial \alpha} = \frac{1}{\alpha} \sum_{t=m+1}^n \left[\frac{4}{\alpha^2} \sinh^2 \left(\frac{y_t - \mu_t}{2} \right) - 1 \right].$$

Notice that

$$\frac{\partial \ell_t}{\partial \mu_t} = \tanh \left(\frac{y_t - \mu_t}{2} \right) \left[\frac{2}{\alpha^2} \cosh^2 \left(\frac{y_t - \mu_t}{2} \right) - \frac{1}{2} \right]$$

and hence

$$\begin{aligned} \frac{\partial \ell_t}{\partial \beta_i} &= \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \beta_i} = \left(x_{t,i} - \sum_{k=1}^p \phi_k x_{t-k,i} \right) \frac{\partial \ell_t}{\partial \mu_t}, \\ \frac{\partial \ell_t}{\partial \eta} &= \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta} = \frac{\partial \ell_t}{\partial \mu_t}. \end{aligned}$$

Thus, the scores associated with the parameters β_i and η are defined as

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_i} &= \sum_{t=m+1}^n \left(x_{t,i} - \sum_{k=1}^p \phi_k x_{(t-k)i} \right) \tanh \left(\frac{y_t - \mu_t}{2} \right) \left[\frac{2}{\alpha^2} \cosh^2 \left(\frac{y_t - \mu_t}{2} \right) - \frac{1}{2} \right], \quad (10) \\ \frac{\partial \ell}{\partial \eta} &= \sum_{t=m+1}^n \tanh \left(\frac{y_t - \mu_t}{2} \right) \left[\frac{2}{\alpha^2} \cosh^2 \left(\frac{y_t - \mu_t}{2} \right) - \frac{1}{2} \right]. \end{aligned}$$

Then, since

$$\begin{aligned} \frac{\partial \ell_t}{\partial \phi_i} &= \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \phi_i} = (y_{t-i} - \mathbf{x}_{t-i} \boldsymbol{\beta}) \frac{\partial \ell_t}{\partial \mu_t}, \\ \frac{\partial \ell_t}{\partial \theta_j} &= \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \theta_j} = u_{t-j} \frac{\partial \ell_t}{\partial \mu_t}, \end{aligned}$$

we have

$$\begin{aligned}\frac{\partial \ell}{\partial \phi_i} &= \sum_{t=m+1}^n (y_{t-i} - \mathbf{x}_{t-i} \boldsymbol{\beta}) \tanh\left(\frac{y_t - \mu_t}{2}\right) \left[\frac{2}{\alpha^2} \cosh^2\left(\frac{y_t - \mu_t}{2}\right) - \frac{1}{2} \right], \\ \frac{\partial \ell}{\partial \theta_j} &= \sum_{t=m+1}^n u_{t-j} \tanh\left(\frac{y_t - \mu_t}{2}\right) \left[\frac{2}{\alpha^2} \cosh^2\left(\frac{y_t - \mu_t}{2}\right) - \frac{1}{2} \right].\end{aligned}$$

The ML estimate, $\hat{\boldsymbol{\gamma}}$, of the vector of parameters $\boldsymbol{\gamma}$ is obtained through the solution of the system of equations $U_\eta(\boldsymbol{\gamma}) = 0$, $U_{\beta_k}(\boldsymbol{\gamma}) = 0$, for $k = 1, \dots, w$, $U_\alpha(\boldsymbol{\gamma}) = 0$, $U_{\phi_i}(\boldsymbol{\gamma}) = 0$, for $i = 1, \dots, p$, and $U_{\theta_j}(\boldsymbol{\gamma}) = 0$, for $j = 1, \dots, q$, where $U_\eta(\boldsymbol{\gamma}) = \partial \ell / \partial \eta$, $U_{\beta_k}(\boldsymbol{\gamma}) = \partial \ell / \partial \beta_k$, $U_\alpha(\boldsymbol{\gamma}) = \partial \ell / \partial \alpha$, $U_{\phi_i}(\boldsymbol{\gamma}) = \partial \ell / \partial \phi_i$ and $U_{\theta_j}(\boldsymbol{\gamma}) = \partial \ell / \partial \theta_j$. Except for $U_\alpha = 0$, where the ML estimate is given by

$$\hat{\alpha} = \sqrt{\frac{4}{n} \sum_{t=m+1}^n \left[\sinh\left(\frac{y_t - \mu_t}{2}\right) \right]^2}, \quad (11)$$

the normal equations $\mathbf{U}(\boldsymbol{\gamma}) = \partial \ell / \partial \boldsymbol{\gamma} = \mathbf{0}$ do not have analytic solutions, and hence it is necessary to use a non-linear optimization method to maximize the log-likelihood function defined in (9). **The Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm (see Lange, 2001)**, also known as the quantum-quantum BFGS algorithm, is a good choice for the solution of non-linear systems since, in most cases, it can obtain the solution more quickly than other methods. This quasi-Newton method approximates the second derivative matrix, reducing the number of operations per iteration. For more details on numerical maximization methods and the BFGS algorithm, see Nocedal and Wright (1999), Press et al. (1992) and Lange (2001). The BFGS algorithm is implemented in **the R software (version 3.6.3)** by the functions `optim` and `optimx`; see `www.R-project.org` and R Core Team (2018). Results for the Fisher information matrix are provided in Appendix B. **Starting values are required to initiate the parameter estimation. Particularly, initial values for $\hat{\boldsymbol{\beta}}$, $\hat{\eta}$, $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$ are obtained from the R function `arima`, whereas the initial value for $\hat{\alpha}$ is obtained from (11).**

Theorem 2. *For each $v = 1, \dots, k$ fixed, suppose that $x_{tv} > p \max_{i=1, \dots, p} |\phi_i x_{(t-i)v}|$ for each $t = m+1, \dots, n$. If $\alpha, \eta, \boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$ are known such that $0 < \alpha < 2$, then there exists a unique maximum likelihood estimate of the parameter β_v , for each $v = 1, \dots, k$. See proof in Appendix A.*

3.3 Prediction

The prediction using origin t and horizon h is denoted by \hat{y}_{t+h} . Consider that

$$\hat{y}_{t+h} = \begin{cases} \hat{y}_t(h), & \text{for } h > 0; \\ y_{t+h}, & \text{for } h \leq 0; \end{cases} \quad \hat{\varepsilon}_{t+h} = \begin{cases} 0, & \text{for } h > 0; \\ \hat{\varepsilon}_{t+h}, & \text{for } h \leq 0. \end{cases}$$

The fact that $\hat{\varepsilon}_{t+h} = 0$ for $h > 0$ indicates that observation y_{t+h} is predicted correctly. Estimates for μ_t , with $t = m+1, \dots, n$, denoted by $\hat{\mu}_t$, are obtained from the ML estimate of $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\gamma}}$, as

$$\hat{\mu}_t = \hat{\eta} + \mathbf{x}_t^\top \hat{\boldsymbol{\beta}} + \sum_{i=1}^p \hat{\phi}_i (y_{t-i} - \mathbf{x}_{t-i}^\top \hat{\boldsymbol{\beta}}) + \sum_{j=1}^q \hat{\theta}_j \hat{u}_{t-j}.$$

Using $\widehat{\mu}_t$, we can get $\widehat{\varepsilon}_t$, with $t = m + 1, \dots, n$. For example, for $\varepsilon_t = y_t - \mu_t$, that is, measurement errors on the original scale, we have $\widehat{\varepsilon}_t = y_t - \widehat{\mu}_t$. Thus, we can predict y_{n+1} by

$$\widehat{y}_{n+1} = \widehat{\eta} + \mathbf{x}_{n+1}^\top \widehat{\boldsymbol{\beta}} + \sum_{i=1}^p \widehat{\phi}_i(y_{n+1-i} - \mathbf{x}_{n+1-i}^\top \widehat{\boldsymbol{\beta}}) + \sum_{j=1}^q \widehat{\theta}_j \widehat{u}_{n+1-j}.$$

For a time $n + 2$, we obtain

$$\widehat{y}_{n+2} = \widehat{\eta} + \mathbf{x}_{n+2}^\top \widehat{\boldsymbol{\beta}} + \sum_{i=1}^p \widehat{\phi}_i(y_{n+2-i} - \mathbf{x}_{n+2-i}^\top \widehat{\boldsymbol{\beta}}) + \sum_{j=1}^q \widehat{\theta}_j \widehat{u}_{n+2-j},$$

and so on.

3.4 Residual analysis

The analysis of residuals plays a fundamental role in the validation of any statistical model. This analysis aims at detection of possible outliers based on an assessment of their effect on the model fitting and prediction. To evaluate the fit of the BS log-linear regression model defined in (3), we consider the generalized Cox-Snell (GCS) residual defined as

$$r_i = -\log(\widehat{S}(y_i | \mathcal{F}_{i-1})), \quad i = m + 1, \dots, n, \quad (12)$$

where \widehat{S} is the survival function estimated for the model evaluated for observation i given by

$$\widehat{S}(Y_i; \alpha, \boldsymbol{\beta}, \mathbf{x}) = \Phi \left[-\frac{2}{\widehat{\alpha}} \sinh \left(\frac{y_i - \widehat{\mu}_i}{2} \right) \right], \quad i = m + 1, \dots, n.$$

The GCS residuals follow a unit exponential distribution, EXP(1), when the model is specified correctly. Because the GCS residuals have an EXP(1) distribution, a **quantile–quantile (QQ) plot of r_i** , defined in (12), for $i = m + 1, \dots, n$, can be used to evaluate the fit of the model. The interested reader can see details about GCS residuals in [Leiva et al. \(2016\)](#) and references therein.

4 Numerical calculations

We present the results of two Monte Carlo simulation studies for the BISARMA(1,1) model. Results are presented only for $p = 1$ and $q = 1$, since higher order of lags produced similar results. The first study considers the evaluation of the performance of the ML estimators, while the second study assesses whether the model is chosen correctly according to the data generating process. The idea in this second study is to generate simulated data from the BISARMA(1,1) model and check if the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the root mean square error (RMSE) can select the correct model. The comparison is made with the Gaussian ARMA(1,1) model. Therefore, it is expected that the AIC, BIC and RMSE measures will indicate that the BISARMA(1,1) model is better.

4.1 Maximum likelihood estimation

In this section, Monte Carlo simulations are used to evaluate the performance and behaviour of the ML estimators of the model parameters. All the simulation and estimation routines are developed using the R software. For the study, in all cases and for each sample size (n), the number of Monte Carlo replications is set at $N_r = 10,000$. The ML estimates are obtained by maximizing the log-likelihood function defined in (9) using the BFGS optimization algorithm. To assess accuracy of estimators, for each case and sample size, the mean, bias, variance and mean square error (MSE) are calculated empirically based on the simulated data, with definitions given, respectively, by

$$\bar{\hat{\varphi}} = \frac{1}{N_r} \sum_{r=1}^{N_r} \hat{\varphi}_r, \quad \text{Bias}(\hat{\varphi}) = \bar{\hat{\varphi}} - \varphi, \quad \widehat{\text{Var}}(\hat{\varphi}) = \frac{1}{N_r} \sum_{r=1}^{N_r} (\hat{\varphi}_r - \bar{\hat{\varphi}})^2, \quad \text{and} \quad \widehat{\text{MSE}}(\hat{\varphi}) = \frac{1}{N_r} \sum_{r=1}^{N_r} (\hat{\varphi}_r - \varphi)^2,$$

where $\hat{\varphi}_r$ is the estimate obtained in the r -th replicate, φ is the true value of the parameter, and N_r is the number of Monte Carlo replicates. With the exception of the bias, for all other statistics, as the value decreases, estimation performance improves – bias only has this result, when analysed in terms of its absolute value.

The results of the simulation study performed using the BISARMA(1,1) model are now presented. We evaluate the performance of the ML estimator for the shape parameter, α , autoregressive parameter, ϕ , and moving average parameter, θ . Specifically, the study is conducted for two cases. In the first case, we consider sample sizes $n \in \{50, 100, 200, 500\}$, with the following values for the parameters of interest: $\alpha \in \{0.25, 0.5, 1.5, 2.5\}$, $\beta = 0.7$, $\eta = 1.0$, $\phi = 0.7$ and $\theta = 0.5$. The sample sizes considered allow us to verify whether there are improvements in the estimation of the model parameters as the sample size increases. The set of values for α is chosen in order to obtain different shapes. In the second case, sample sizes $n \in \{50, 100, 200, 500\}$, with values of $\alpha = 0.5$, $\beta = 0.7$ and $\eta = 1.0$ are considered. The parameters ϕ and θ are fixed at values of 0.3, 0.5 and 0.7. The results of the calculated statistics for the first and second case are reported in Tables 1 and 2.

The results in Table 1 show that, in general, the performance of the estimates of α is directly related to the sample size. That is, as n increases, the accuracy of the estimates improves, as expected. These conclusions are also valid for α , β , η , ϕ and θ . Note that when the sample size increases from $n = 50$ to $n = 500$, the bias in absolute value of the estimator of $\alpha = 0.5$, on average, decreases considerably, going from 0.0303 to 0.0028. For a fixed sample size, the absolute bias of the estimators increases as α increases. For example, when $n = 100$ and $\alpha = 0.25$, the absolute bias of $\hat{\alpha}$ is 0.0071. However, when $\alpha = 2.5$, this bias increases to 0.0813. Such a behaviour is similar for the variance and MSE of the estimator of α . In all scenarios considered, the parameter α is, on average, underestimated, that is, $\hat{\alpha}$ is less than the true value of the parameter. Note that the variability of $\hat{\alpha}$ decreases when n increases, as expected, with similar results for the MSE. This indicates that the estimator of the parameter α of the BISARMA(1,1) model obtained by the ML estimator is accurate.

Figure 3 shows the results in graphical form for the indicated values of n and α . Note in Figure 3(a) that, as n increases, the bias of the estimator is smaller in absolute value. Also for a fixed sample size, the absolute bias increases as α increases. This behaviour is similar for the variance and MSE of $\hat{\alpha}$; see Figures 3(b) and (c).

For the second case, Table 2 reports summary statistics for the estimates of the parameters ϕ and θ . Note that the ML estimator for ϕ and θ are accurate. For example, for $n = 500$, $\phi = 0.5$ and $\theta = 0.3$, the estimates are quite close to the true value of the parameters, that is, $\hat{\phi} = 0.4914$ and $\hat{\theta} = 0.3067$. On average, the empirical absolute bias of $\hat{\phi}$ and $\hat{\theta}$ are always less than 0.0490. The largest values

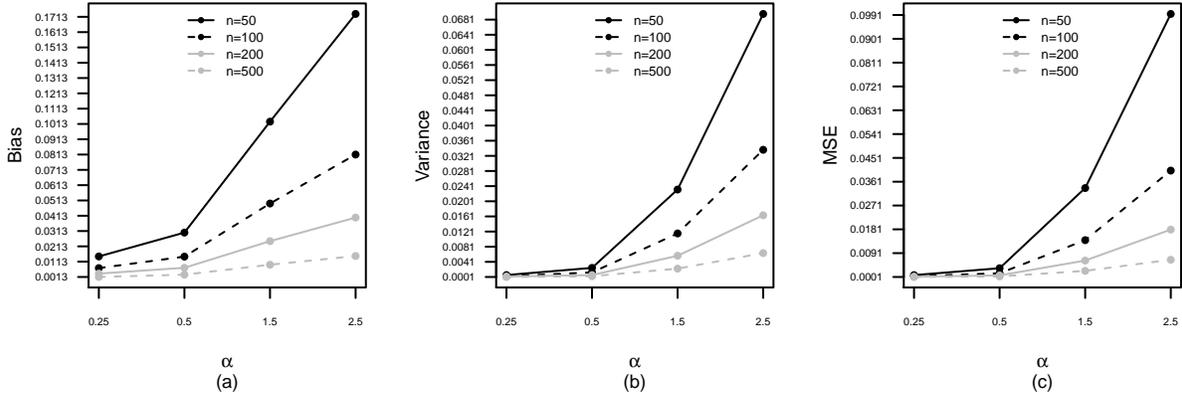


Figure 3: Empirical absolute bias (a), variance (b) and MSE (c) of the estimator of α for simulated data.

of the empirical MSEs occur when $\phi = 0.3$, $\theta = 0.3$, and $n = 100$. It can also be seen that the ML estimators of ϕ and θ become more accurate as the sample size increases. Considering a fixed sample size, we note a slight reduction in the variance and MSE of $\hat{\phi}$ and $\hat{\theta}$, as the values of ϕ and/or θ increases. Also note that ϕ is, on average, underestimated, that is, $\hat{\phi}$ is less than its true value, in all scenarios considered, whereas θ is overestimated.

Table 1: ML estimates for indicated α , based on Monte Carlo simulations for the BISARMA(1,1) model.

n	α	$\hat{\alpha}$			
		Mean	Bias	Variance	MSE
50	0.25	0.2353	-0.0147	0.0006	0.0008
	0.5	0.4697	-0.0303	0.0025	0.0034
	1.5	1.3972	-0.1028	0.0232	0.0337
	2.5	2.3268	-0.1732	0.0696	0.0995
100	0.25	0.2429	-0.0071	0.0003	0.0004
	0.5	0.4854	-0.0146	0.0013	0.0015
	1.5	1.4507	-0.0493	0.0116	0.0140
	2.5	2.4187	-0.0813	0.0337	0.0403
200	0.25	0.2465	-0.0035	0.0002	0.0002
	0.5	0.4927	-0.0073	0.0006	0.0007
	1.5	1.4753	-0.0247	0.0057	0.0063
	2.5	2.4599	-0.0401	0.0164	0.0180
500	0.25	0.2487	-0.0013	0.0001	0.0001
	0.5	0.4972	-0.0028	0.0003	0.0003
	1.5	1.4907	-0.0093	0.0023	0.0024
	2.5	2.4850	-0.0150	0.0064	0.0066

Table 2: ML estimates for the different values of ϕ and θ , based on Monte Carlo simulations for the BIS-ARMA(1,1) model.

n	ϕ	θ	$\hat{\phi}$				$\hat{\theta}$			
			Mean	Bias	Variance	MSE	Mean	Bias	Variance	MSE
50	0.3	0.3	0.1893	-0.1107	0.0839	0.0961	0.4094	0.1094	0.0869	0.0989
		0.5	0.2087	-0.0913	0.0546	0.0629	0.5931	0.0931	0.0516	0.0603
		0.7	0.2230	-0.0770	0.0397	0.0456	0.7790	0.0790	0.0368	0.0430
	0.5	0.3	0.3991	-0.1009	0.0463	0.0564	0.3950	0.0950	0.0546	0.0636
		0.5	0.4112	-0.0888	0.0346	0.0424	0.5870	0.0870	0.0389	0.0464
		0.7	0.4205	-0.0795	0.0278	0.0341	0.7779	0.0779	0.0320	0.0380
	0.7	0.3	0.6023	-0.0977	0.0259	0.0354	0.3868	0.0868	0.0387	0.0462
		0.5	0.6093	-0.0907	0.0215	0.0297	0.5840	0.0840	0.0321	0.0391
		0.7	0.6138	-0.0862	0.0186	0.0260	0.7825	0.0825	0.0283	0.0351
100	0.3	0.3	0.2510	-0.0490	0.0364	0.0388	0.3435	0.0435	0.0377	0.0396
		0.5	0.2584	-0.0416	0.0243	0.0261	0.5383	0.0383	0.0236	0.0250
		0.7	0.2644	-0.0356	0.0185	0.0198	0.7311	0.0311	0.0162	0.0172
	0.5	0.3	0.4562	-0.0438	0.0188	0.0207	0.3367	0.0367	0.0233	0.0246
		0.5	0.4595	-0.0405	0.0149	0.0165	0.5354	0.0354	0.0178	0.0190
		0.7	0.4632	-0.0368	0.0124	0.0138	0.7306	0.0306	0.0140	0.0149
	0.7	0.3	0.6575	-0.0425	0.0097	0.0115	0.3334	0.0334	0.0166	0.0177
		0.5	0.6587	-0.0413	0.0086	0.0103	0.5343	0.0343	0.0145	0.0156
		0.7	0.6607	-0.0393	0.0076	0.0091	0.7322	0.0322	0.0126	0.0136
200	0.3	0.3	0.2768	-0.0232	0.0170	0.0175	0.3198	0.0198	0.0177	0.0181
		0.5	0.2789	-0.0211	0.0115	0.0119	0.5182	0.0182	0.0110	0.0113
		0.7	0.2828	-0.0172	0.0090	0.0093	0.7145	0.0145	0.0078	0.0080
	0.5	0.3	0.4794	-0.0206	0.0087	0.0091	0.3165	0.0165	0.0111	0.0113
		0.5	0.4800	-0.0200	0.0070	0.0074	0.5163	0.0163	0.0084	0.0087
		0.7	0.4823	-0.0177	0.0060	0.0063	0.7141	0.0141	0.0068	0.0070
	0.7	0.3	0.6799	-0.0201	0.0043	0.0047	0.3150	0.0150	0.0078	0.0081
		0.5	0.6800	-0.0200	0.0038	0.0042	0.5155	0.0155	0.0069	0.0071
		0.7	0.6812	-0.0188	0.0035	0.0039	0.7150	0.0150	0.0061	0.0063
500	0.3	0.3	0.2904	-0.0096	0.0066	0.0067	0.3080	0.0080	0.0069	0.0070
		0.5	0.2916	-0.0084	0.0045	0.0046	0.5073	0.0073	0.0043	0.0044
		0.7	0.2923	-0.0077	0.0035	0.0036	0.7062	0.0062	0.0031	0.0031
	0.5	0.3	0.4914	-0.0086	0.0034	0.0035	0.3067	0.0067	0.0043	0.0044
		0.5	0.4920	-0.0080	0.0027	0.0028	0.5066	0.0066	0.0033	0.0033
		0.7	0.4922	-0.0078	0.0023	0.0024	0.7060	0.0060	0.0026	0.0027
	0.7	0.3	0.6917	-0.0083	0.0016	0.0017	0.3061	0.0061	0.0030	0.0031
		0.5	0.6919	-0.0081	0.0015	0.0015	0.5064	0.0064	0.0026	0.0027
		0.7	0.6919	-0.0081	0.0014	0.0014	0.7061	0.0061	0.0023	0.0024

4.2 Performance measures and model selection

A fundamental step in statistical modeling is selecting a specific model which describes the data well, and which is also valid for prediction. Several performance measures can be used to assess the accuracy of forecasts and to compare different models. The most widely used measure of goodness of fit is the **RMSE** defined by

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2},$$

where n is the length of the series, y_t is the observed value of Y_t at t and \hat{y}_t is its predicted value at t . Whereas for model selections, the **AIC** and **BIC** are preferred, and are defined as

$$\text{AIC} = -2\log(L) + 2k, \quad \text{and} \quad \text{BIC} = -2\log(L) + 2k\log(n),$$

where L is the maximized value of the likelihood function for the estimated model, n is the number of observations in the sample, and k is the number of estimated parameters. The model that provides the minimum of AIC or BIC is selected as the best fit model. Table 3 presents results of these measures for sample sizes $n \in \{50, 100, 200, 500\}$ for the **BISARMA(1,1)** model, with $\eta = 1.0$, $\beta = 0.7$, $\alpha = 0.5$, $\phi \in \{0.3, 0.5, 0.7\}$ and $\theta \in \{0.3, 0.5, 0.7\}$. For each combination of parameters, 1,000 Monte Carlo replicates are used. Comparing the **BISARMA(1,1)** and Gaussian **ARMA(1,1)** models based on the statistics given in Table 3, it is confirmed that the AIC and/or BIC values highlight the fact that the **BISARMA** model fits the data better than the Gaussian **ARMA** model. The **BISARMA(1,1)** model also provides smaller RMSE values, indicating better prediction ability. To verify the effects of the shape parameter α on the model performance, we consider the values of $\eta = 1.0$, $\beta = 0.7$, $\phi = 0.6$, $\theta = 0.2$ and $\alpha \in \{0.5, 1.0, 1.5, 2.0, 2.5\}$, again with 1,000 Monte Carlo replicates. The results are reported in Table 4. In this case, the **BISARMA(1,1)** model again provides smaller values of AIC, BIC and RMSE, indicating a better performance than the Gaussian **ARMA(1,1)** model.

4.3 Analytics of real-world mortality time series data

Next, three real time-series data sets are used to illustrate the performance of the **BISARMA(p,q)** model. The data correspond to time series with $n = 508$ continuous observations. They refer to the possible effects of temperature variation and levels of particulate matter (particulates) on cardiovascular mortality in Los Angeles between 1970 and 1979; see Figure 4. Data are available in the R software through the package `astsa`; see Shumway and Stoffer (2017). The variables under study are mortality (M_t), temperature (X_{1t}), and particulates (X_{2t}).

Figure 4(a) shows a downward trend along the series M_t . In all series, the presence of seasonal peaks, corresponding to winter-summer variations, is also evident. Figure 5 shows the dispersion plots for the relationship between: (a) mortality and particulates and (b) mortality and temperature. Figure 5(a) indicates a possible linear relationship between mortality and particulates levels whereas Figure 5(b) shows a curvilinear relationship between mortality and temperature, indicating that mortality fluctuations are strongly associated with temperature variations. **Shumway and Stoffer (2017) used these data in several examples throughout their book. Amongst these examples, we can highlight the classical multiple linear regression and ARMA applications.** Figure 5 shows the relationship between variables M_t , X_{1t} and X_{2t} . Considering M_t as the response variable, these relationships can

Table 3: Predictive evaluation statistics for the different values of ϕ and θ , based on Monte Carlo simulations for the BISARMA (1,1) model, with statistics for the ARMA (1,1) model in parentheses.

n	ϕ	θ	AIC	BIC	RMSE
50		0.3	72.3576 (73.3029)	81.9177 (82.8631)	0.4603 (0.4770)
	0.3	0.5	71.9333 (73.9673)	81.4934 (83.5274)	0.4585 (0.4824)
		0.7	72.2643 (75.5863)	81.8245 (85.1464)	0.4599 (0.4959)
	0.5	0.3	72.2507 (73.3638)	81.8108 (82.9240)	0.4598 (0.4786)
		0.5	71.8260 (74.0690)	81.3862 (83.6292)	0.4580 (0.4849)
		0.7	72.2174 (75.9380)	81.7775 (85.4981)	0.4597 (0.5015)
	0.7	0.3	72.0982 (73.3189)	81.6584 (82.8790)	0.4591 (0.4801)
		0.5	71.6726 (74.1032)	81.2327 (83.6633)	0.4572 (0.4875)
	0.7	72.0363 (76.2311)	81.5964 (85.7912)	0.4588 (0.5073)	
100		0.3	141.8991 (142.7913)	154.9250 (155.8172)	0.4729 (0.4814)
	0.3	0.5	141.9712 (143.7678)	154.9971 (156.7937)	0.4731 (0.4845)
		0.7	141.9093 (145.7418)	154.9351 (158.7677)	0.4728 (0.4915)
	0.5	0.3	141.7895 (142.8292)	154.8154 (155.8551)	0.4726 (0.4822)
		0.5	141.8908 (143.9018)	154.9166 (156.9276)	0.4729 (0.4860)
		0.7	141.8989 (146.1938)	154.9248 (159.2196)	0.4727 (0.4945)
	0.7	0.3	141.6287 (142.8097)	154.6546 (155.8356)	0.4722 (0.4830)
		0.5	141.7572 (143.9976)	154.7831 (157.0235)	0.4726 (0.4876)
	0.7	141.8046 (146.6471)	154.8304 (159.6729)	0.4725 (0.4977)	
200		0.3	280.6618 (281.7446)	297.1533 (298.2362)	0.4787 (0.4829)
	0.3	0.5	280.8021 (282.8071)	297.2937 (299.2987)	0.4789 (0.4842)
		0.7	280.8887 (285.1818)	297.3803 (301.6734)	0.4790 (0.4888)
	0.5	0.3	280.5644 (281.7629)	297.0559 (298.2545)	0.4786 (0.4833)
		0.5	280.7148 (282.9062)	297.2064 (299.3978)	0.4788 (0.4848)
		0.7	280.9012 (285.7727)	297.3928 (302.2643)	0.4790 (0.4905)
	0.7	0.3	280.4677 (281.7850)	296.9593 (298.2765)	0.4785 (0.4839)
		0.5	280.6136 (283.0111)	297.1051 (299.5027)	0.4787 (0.4856)
	0.7	280.8398 (286.4212)	297.3314 (302.9128)	0.4789 (0.4924)	
500		0.3	699.6953 (701.3716)	720.7683 (722.4447)	0.4837 (0.4854)
	0.3	0.5	700.0944 (702.7698)	721.1674 (723.8428)	0.4839 (0.4863)
		0.7	699.3433 (704.6478)	720.4163 (725.7208)	0.4835 (0.4875)
	0.5	0.3	699.5842 (701.3973)	720.6573 (722.4703)	0.4836 (0.4855)
		0.5	699.9943 (702.8966)	721.0674 (723.9697)	0.4838 (0.4866)
		0.7	699.2964 (705.2882)	720.3694 (726.3612)	0.4835 (0.4882)
	0.7	0.3	699.4819 (701.4246)	720.5549 (722.4977)	0.4836 (0.4857)
		0.5	699.9119 (703.0665)	720.9849 (724.1395)	0.4838 (0.4869)
	0.7	699.1754 (706.0010)	720.2485 (727.0740)	0.4834 (0.4890)	

Table 4: Predictive evaluation statistics for the different values of α , based on Monte Carlo simulations for the BISARMA (1,1) model, with statistics for the ARMA (1,1) model in parentheses.

n	α	AIC	BIC	RMSE
50	0.5	71.8284 (72.6513)	81.3885 (82.2114)	0.4581 (0.4768)
	1.0	133.2322 (134.7363)	142.7924 (144.2964)	0.8646 (0.8821)
	1.5	164.9856 (167.7778)	174.5457 (177.3379)	1.2151 (1.2306)
	2.0	184.8822 (189.2896)	194.4424 (198.8497)	1.5198 (1.5298)
	2.5	198.6240 (204.8004)	208.1841 (214.3605)	1.7880 (1.7903)
100	0.5	141.9090 (142.8055)	154.9349 (155.8314)	0.4729 (0.4813)
	1.0	265.7652 (268.0712)	278.7910 (281.097)	0.8913 (0.8985)
	1.5	329.6598 (334.6509)	342.6856 (347.6767)	1.2499 (1.2553)
	2.0	369.6198 (378.0031)	382.6457 (391.0290)	1.5593 (1.5614)
	2.5	397.1729 (409.2500)	410.1987 (422.2758)	1.8297 (1.8275)
200	0.5	280.5448 (281.5542)	297.0364 (298.0457)	0.4786 (0.4831)
	1.0	-349.1334 (-347.1097)	-332.6419 (-330.6182)	0.0983 (0.1120)
	1.5	657.9656 (666.9540)	674.4572 (683.4456)	1.2625 (1.2651)
	2.0	738.2936 (753.9990)	754.7852 (770.4906)	1.5730 (1.5737)
	2.5	793.6655 (816.7358)	810.1571 (833.2274)	1.8436 (1.8419)
500	0.5	699.4671 (701.0986)	720.5401 (722.1716)	0.4836 (0.4852)
	1.0	1323.3460 (1331.902)	1344.4190 (1352.9750)	0.9101 (0.9113)
	1.5	1644.9330 (1666.830)	1666.0060 (1687.9030)	1.2736 (1.2743)
	2.0	1845.8550 (1884.830)	1866.9280 (1905.9030)	1.5853 (1.5852)
	2.5	1984.2180 (2041.935)	2005.2910 (2063.0080)	1.8563 (1.8552)

be modelled over time t by means of the observed values of X_{1t} and X_{2t} , x_{1t} and x_{2t} say, as

$$M_t = \eta + \beta_1 \text{trend}_t + \beta_2(x_{1t} - \bar{x}_1) + \beta_3(x_{1t} - \bar{x}_1)^2 + \beta_4 x_{2t} + \varepsilon_t, \quad (13)$$

where \bar{x}_1 is the mean temperature added in the model to avoid collinearity, “trend $_t$ ” is the downward linear trend observed in Figure 4(a), and ε_t is a random error, or a noise process, consisting of independent variables which are identically normal distributed with zero mean and variance σ_ε^2 ; see Shumway and Stoffer (2017). In time series regression modeling, it is unusual for the noise to be white, and eventually this assumption must be relaxed. Figure 6 shows the autocorrelation function (ACF) in (a) and partial autocorrelation function (PACF) in (b) of the residuals from the least squares fit of (13). The ACF for an AR(p) process slowly tends to zero and the PACF cuts off at lag p ; see Table 3.1 of Shumway and Stoffer (2017). Therefore, examination of the ACF and PACF in Figure 6 suggests a stationary AR(p) model of order $p = 2$ for the residuals. Then, the correlated error model defined in (13) can be expressed as $\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + u_t$, where ε_t is an AR(2) model and u_t is a white noise process. The results for this model are obtained using the function `gammaFit` of the package `gamlss.util`; see <http://www.gamlss.org>.

Consider now an analysis of the **BISARMA(2, 0)** model defined in (7), with

$$Y_t = \log(M_t) = \eta + \beta_1 \text{trend}_t + \beta_2(x_{1t} - \bar{x}_1) + \beta_3(x_{1t} - \bar{x}_1)^2 + \beta_4 x_{2t} + \sum_{i=1}^2 \phi_i [y_{t-i} - (\eta + \beta_1 \text{trend}_{t-i} + \beta_2(x_{1t-i} - \bar{x}_1) + \beta_3(x_{1t-i} - \bar{x}_1)^2 + \beta_4 x_{2t-i})] + \varepsilon_t,$$

where $\varepsilon_t \sim \text{log-BS}(\alpha, 0)$, for the data set associated with the variables M_t (with $M_t = T_t$ according to (8)), X_{1t} and X_{2t} . Table 5 reports the values of the ML estimates, RMS, AIC and BIC. This table also reports the p -values of the Ljung-Box (LB) statistic, labelled $Q(k)$, for up to k -th order serial correlation. This statistic evaluates the autocorrelation of the GCS residuals. From this table, note that the BISARMA(2,0) model provides a better fit than the Gaussian ARMA(2,0) model based on the AIC and BIC values. Also, the BISARMA(2,0) model has smaller RMSE values, indicating better quality of the model. Finally, the LB p -values provide no evidence of serial correlation in the GCS residuals for the two models. Nevertheless, those LB p -values favour the BISARMA(2,0) model.

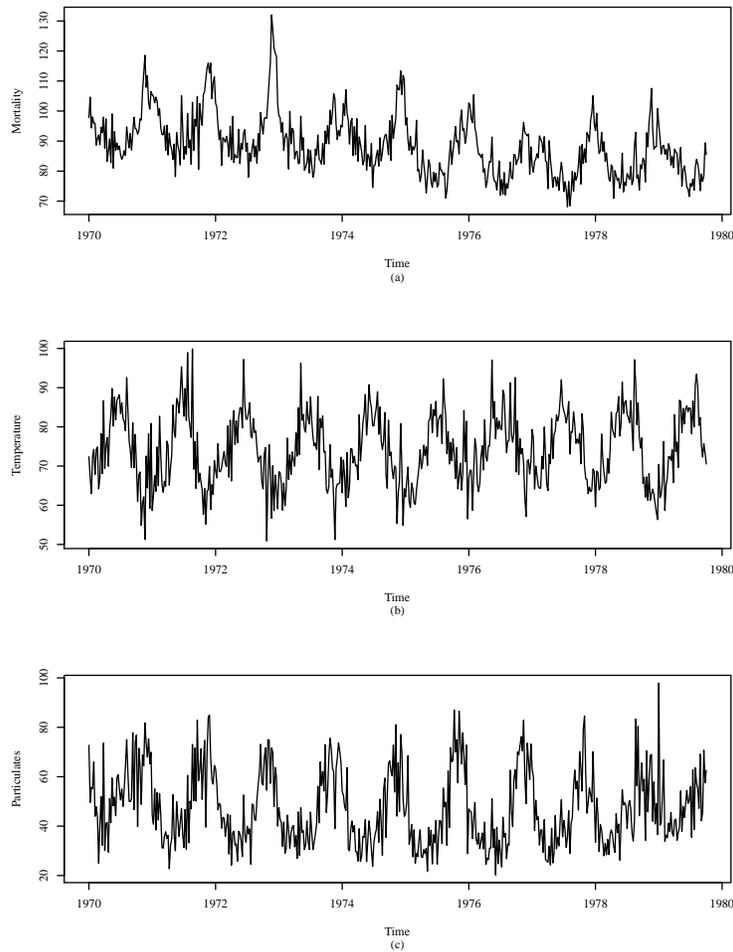


Figure 4: Cardiovascular mortality (a), temperature (b) and particulates (c) over the 10-year period (1970 - 1979) in Los Angeles.

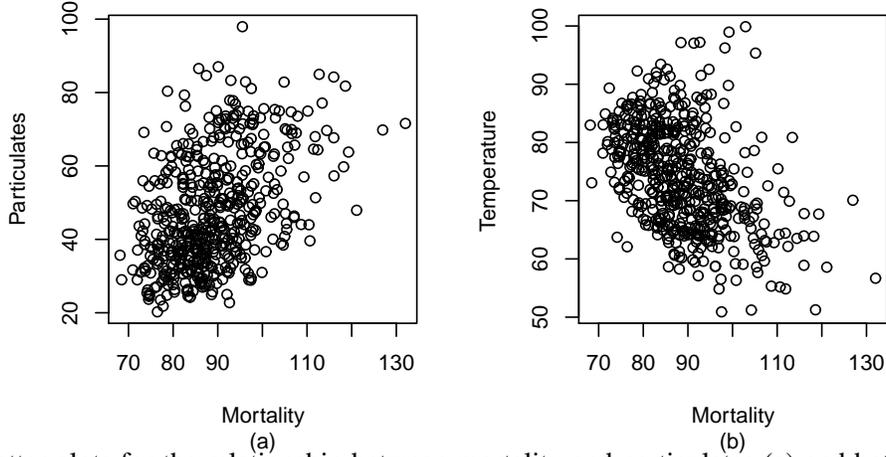


Figure 5: Scatter plots for the relationship between mortality and particulates (a) and between mortality and temperature (b) over the 10-year period (1970 - 1979) in Los Angeles.

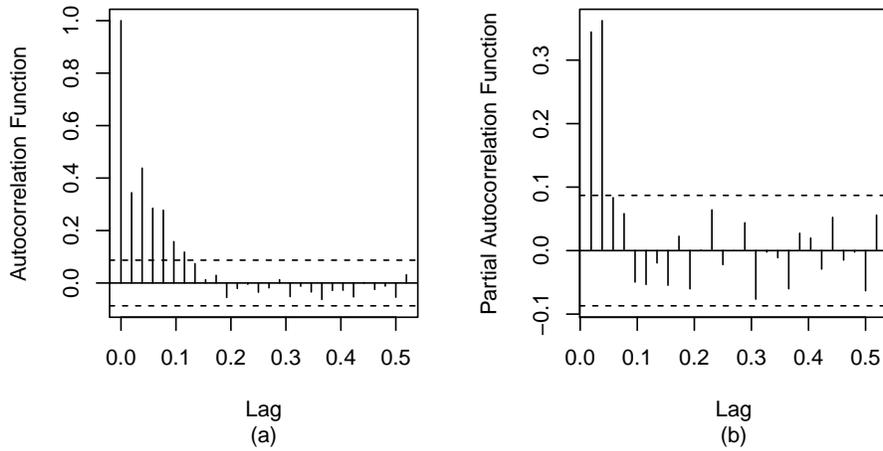


Figure 6: plots of ACF (a) and PACF (b) of the residuals of the regression model for times series data over the 10-year period (1970 - 1979) in Los Angeles.

The QQ plots of the GCS residuals, with simulation envelopes, indicate better agreement with the EXP(1) distribution in the BISARMA(2,0) model; see Figure 7 (left) and Figure 8 (left). From the ACF and PACF plots, however, note that both models produce non-autocorrelated **GCS residuals**; see Figures 7 and 8 (centre and right). The time series fitted by the BISARMA(2,0) and Gaussian ARMA(2,0) models are presented together with the observed time series in Figure 9.

5 Conclusions and future research

The Birnbaum-Saunders distribution is studied often in the statistical literature due to its versatility and efficiency in diverse applications, but few works have considered data with a dependency structure. To fill this gap, in this work we proposed a novel autoregressive moving average model based on the Birnbaum-Saunders distribution, which allows modelling of non-negative and asymmetric data that have a structure of dependence over time. The new model has proven to be a good alternative to

Table 5: Estimates and **model selection measures** for the mortality times series data over the 10-year period (1970 - 1979) in Los Angeles.

Model	Parameters	Estimate	AIC	BIC	RMSE	Q(4)	Q(16)
BISARMA(2,0)	ϕ_1	0.4057					
	ϕ_2	0.2779					
	η	38.066					
	β_1	-0.0171	-1487.9150	-1454.0710	0.0547	0.9143	0.6037
	β_2	-0.0017					
	β_3	0.0002					
	β_4	0.0023					
	α	0.0548					
ARMA(2,0)	ϕ_1	0.3727					
	ϕ_2	0.4434					
	η	36.4848					
	β_1	-0.0163	-1453.30	-1419.45	0.2890	0.4276	0.2290
	β_2	-0.00003					
	β_3	0.0002					
	β_4	0.0017					

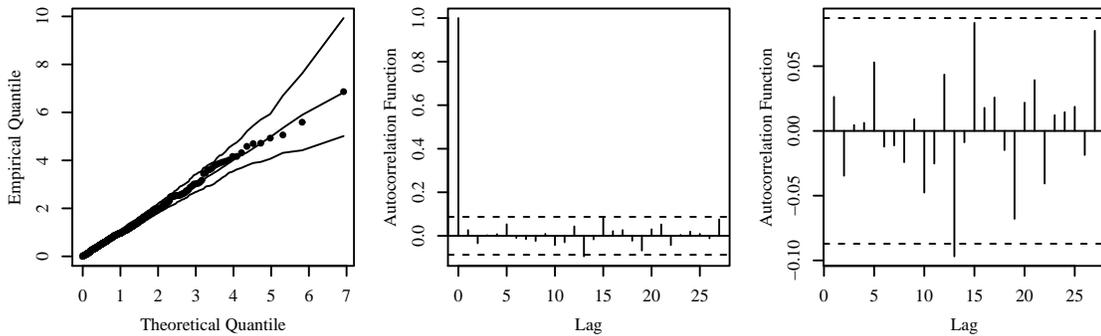


Figure 7: Envelope plot of generalized Cox-Snell residuals (left) and ACF (centre) and PACF (right) for the BISARMA(2,0) model with times series data over the 10-year period (1970 - 1979) in Los Angeles.

describe this type of data with this structure. Some further examples where our model can be applied are in the description of: (i) index of coal production in industry (Rahul et al., 2018); (ii) in chemical process concentrations (McLeod and Zhang, 2006); (iii) river flow (Sim, 1987); and (iv) monthly average affluent streamflow and monthly averages series of the bus ridership (Milani et al., 2017). We performed a numerical evaluation of the maximum likelihood estimators of the model parameters through Monte Carlo simulations using as performance measures the empirical mean, bias, variance and mean square error of the estimators. The simulation study demonstrated the good performance of the maximum likelihood estimators. A diagnostic analysis based on residuals was also carried out. An application of the new Birnbaum-Saunders autoregressive moving average model was performed

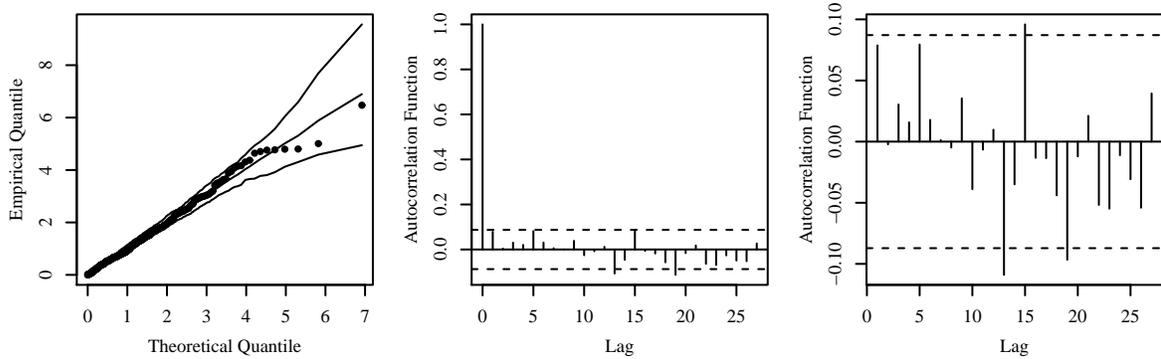


Figure 8: Envelope plot of QQ residuals (left) and autocorrelation (centre) and partial autocorrelation (right) for an ARMA(2,0) model with times series data over the 10-year period (1970 - 1979) in Los Angeles.

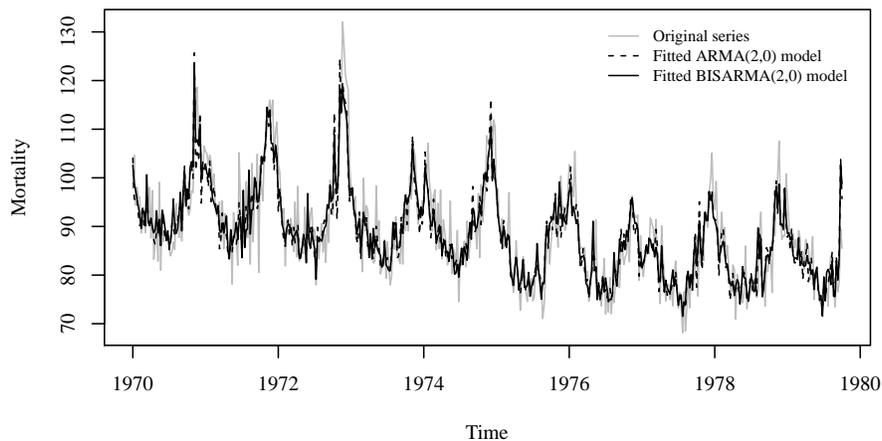


Figure 9: Series cardiovascular mortality in Los Angeles (gray), adjusted by a BISARMA(2,0) model (black) and adjusted by an ARMA(2,0) model (black - - -).

with real-world data from the environmental sciences. The application showed the superiority of the new model over the standard Gaussian moving average autoregressive model, providing strong evidence that the Birnbaum-Saunders distribution is a good modeling alternative when dealing with temporal data. These results suggest that the BISARMA model can become a new standard for the routine analysis of non-negative and asymmetric time series data in the environmental sciences, and elsewhere.

Since the Birnbaum-Saunders distribution is based on the normal distribution, parameter estimation in BISARMA models can be affected by atypical cases. To achieve robust estimation, however, the Birnbaum-Saunders-t distribution model, for example, can be considered instead; see Athayde et al. (2019). Second, in addition to fixed effects, added to a regression model, random effects can also be added to produce mixed models, which may lead to a more sophisticated BISARMA model and more closely describing reality; see Villegas et al. (2011). Third, we can consider multivariate

BISARMA modeling base on the multivariate Birnbaum-Saunders distribution; see Marchant et al. (2018). Fourth, local influence diagnostics can be conducted, which permit detection of combined influential cases. Work on local influence in Birnbaum-Saunders models was conducted by a number of authors; see, for example, Garcia-Papani et al. (2017, 2018b), Santana et al. (2011), Desousa et al. (2018), and Saulo et al. (2019). Fifthly, a simulation study to evaluate the adequacy of generalized Cox-Snell residuals is being considered by the authors (see Leiva et al., 2016), which, unfortunately, in complex models, such as that proposed in this work, may be very intensive computationally. Research on these issues is in progress and their findings will be reported in a future article.

Data availability statement

The data set is available upon request from the authors.

Acknowledgements

The authors would like to thank the editors and reviewers very much for their constructive comments on an earlier version of this manuscript which resulted in this improved version. The research was supported partially by grant “Fondecyt 1200525” from the National Commission for Scientific and Technological Research of the Chilean government.

References

- Athayde, E., Azevedo, A., Barros, M. and Leiva, V. (2019). Failure rate of Birnbaum-Saunders distributions: shape, change-point, estimation and robustness. *Brazilian Journal of Probability and Statistics*, 33:301–328.
- Aykroyd, R.G., Leiva, V., and Marchant, C. (2018). Multivariate Birnbaum-Saunders distributions: Modelling and applications. *Risks* 6(1), article 21 pages:1-25.
- Balakrishnan, N. and Kundu, D. (2019). Birnbaum-Saunders distribution: A review of models, analysis, and application. *Applied Stochastic Models in Business and Industry*, 35:4-49.
- Ben Amor, S., Boubaker, H., and Belkacem, L. (2018). Forecasting electricity spot price for Nord Pool market with a hybrid k-factor GARMA-LLWNN model. *Journal of Forecasting*, 37:832–851.
- Benjamin, M.A., Rigby, R.A., and Stasinopoulos, D.M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98:214–223.
- Bhatti, C. (2010). The Birnbaum-Saunders autoregressive conditional duration model. *Mathematics and Computers in Simulation*, 80:2062–2078.
- Birnbaum, Z.W. and Saunders, S.C. (1969). A new family of life distributions. *Journal of Applied Probability*, 6:319–327.
- Desousa, M.F., Saulo, H., Leiva, V., and Scalco, P. (2018). On a tobit-Birnbaum-Saunders model with an application to antibody response to vaccine. *Journal of Applied Statistics*, 45:932–955.
- Engle, R. and Russell, J. (1998). Autoregressive conditional duration: A new method for irregularly spaced transaction data. *Econometrica*, 66:1127–1162.
- Ferreira, M., Gomes, M.I., and Leiva, V. (2012). On an extreme value version of the Birnbaum-Saunders distribution. *REVSTAT Statistical Journal*, 10:181–210.
- Fonseca, R.V. and Cribari-Neto, F. (2018). Bimodal Birnbaum-Saunders generalized autoregressive score model. *Journal of Applied Statistics*, 45:2585–2606.

- Garcia-Papani, F., Leiva, V., Ruggeri, F., and Uribe-Opazo, M.A. (2018a). Kriging with external drift in a Birnbaum-Saunders geostatistical model. *Stochastic Environmental Research and Risk Assessment*, 32:1517–1530.
- Garcia-Papani, F., Leiva, V., Uribe-Opazo, M.A., and Aykroyd, R.G. (2018b). Birnbaum-Saunders spatial regression models: Diagnostics and application to chemical data. *Chemometrics and Intelligent Laboratory Systems*, 177:114–128.
- Garcia-Papani, F., Uribe-Opazo, M.A., Leiva, V., and Aykroyd, R.G. (2017). Birnbaum-Saunders spatial modelling and diagnostics applied to agricultural engineering data. *Stochastic Environmental Research and Risk Assessment*, 31:105–124.
- Huerta, M., Leiva, V., Liu, S., Rodriguez, M., and Villegas, D (2019). On a partial least squares regression model for asymmetric data with a chemical application in mining. *Chemometrics and Intelligent Laboratory Systems*, 190:55–68.
- Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 2. Wiley, New York, US.
- Lang, S. (1998). *A First Course in Calculus*. Springer, New York, US.
- Lange, K. (2001). *Numerical Analysis for Statisticians*. Springer, New York, US.
- Leão, J., Leiva, V., Saulo, H., and Tomazella, V. (2017). Birnbaum-Saunders frailty regression models: Diagnostics and application to medical data. *Biometrical Journal*, 59:291–314.
- Leão, J., Leiva, V., Saulo, H., and Tomazella, V. (2018a). A survival model with Birnbaum-Saunders frailty for uncensored and censored cancer data. *Brazilian Journal of Probability and Statistics*, 32:707–729.
- Leão, J., Leiva, V., Saulo, H., and Tomazella, V. (2018b). Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. *Statistics in Medicine*, 37:4421–4440.
- Leiva, V. (2016). *The Birnbaum-Saunders Distribution*. Academic Press, New York, US.
- Leiva, V., Ferreira, M., Gomes, M.I., and Lillo, C. (2016). Extreme value Birnbaum-Saunders regression models applied to environmental data. *Stochastic Environmental Research and Risk Assessment*, 30:1045–1058.
- Leiva, V., Marchant, C., Ruggeri, F., and Saulo, H. (2015). A criterion for environmental assessment using Birnbaum-Saunders attribute control charts. *Environmetrics*, 26:463–476.
- Leiva, V., Saulo, H., Leão, J., and Marchant, C. (2014). A family of autoregressive conditional duration models applied to financial data. *Computational Statistics and Data Analysis*, 79:175–191.
- Lillo, C., Leiva, V., Nicolis, O., and Aykroyd, R.G. (2018). L-moments of the Birnbaum-Saunders distribution and its extreme value version: Estimation, goodness of fit and application to earthquake data. *Journal of Applied Statistics*, 45:187–209.
- Maior, V.Q.S. and Cysneiros, F. J.A. (2016). Symarma: A new dynamic model for temporal data on conditional symmetric distribution. *Statistical Papers*, 59:75–97.
- Marchant, C., Leiva, V., Cavieres, M., and Sanhueza, A. (2013). Air contaminant statistical distributions with application to PM10 in Santiago, Chile. *Reviews of Environmental Contamination and Toxicology*, 223:1–31.
- Marchant, C., Leiva, V., and Cysneiros, F. J.A. (2016a). A multivariate log-linear model for Birnbaum-Saunders distributions. *IEEE Transactions on Reliability*, 65:816–827.
- Marchant, C., Leiva, V., Cysneiros, F. J.A., and Liu, S. (2018). Robust multivariate control charts based on Birnbaum-Saunders distributions. *Journal of Statistical Computation and Simulation*, 88:182–202.
- Marchant, C., Leiva, V., Cysneiros, F. J.A., and Vivanco, J.F. (2016b). Diagnostics in multivariate generalized Birnbaum-Saunders regression models. *Journal of Applied Statistics*, 43:2829–2849.
- Marchant, C., Leiva, V., Christakos, G., and Cavieres, M.F. (2019). Monitoring urban environmental pollution by bivariate control charts: new methodology and case study in Santiago, Chile. *Environmetrics*, 30:e2551.

- Martinez, S., Giraldo, R., and Leiva, V. (2019). Birnbaum-Saunders functional regression models for spatial data. *Stochastic Environmental Research and Risk Assessment*, 33:1765–1780.
- McLeod, A.I. and Zhang, Y. (2006). Partial autocorrelation parameterization for subset autoregression. *Journal of Time Series Analysis*, 27:599–612.
- Milani, E., Andrade, M., and Diniz, C. (2017). Generalized normal ARMA model applied to the areas of economy, hydrology, and public policy. *Communications in Statistics: Simulation And Computation*, 46:5819–5835.
- Nocedal, J. and Wright, S. (1999). *Numerical Optimization*. Springer, New York, US.
- Press, W.H., Teulosky, S.A., Vetterling, W.T., and Flannery, B.P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Prentice-Hall, London, UK.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rahul, T., Balakrishnan, N., and Balakrishna, N. (2018). Time series with Birnbaum-Saunders marginal distributions. *Applied Stochastic Models in Business and Industry*, 34:562–581.
- Rieck, J.R. and Nedelman, J.R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 3:51–60.
- Rocha, A.V. and Cribari-Neto, F. (2009). Beta autoregressive moving average models. *TEST*, 18:529–545.
- Santana, L., Vilca, F., and Leiva, V. (2011). Influence analysis in skew-Birnbaum-Saunders regression models and applications. *Journal of Applied Statistics*, 38:1633–1649.
- Saulo, H., Leão, J., Leiva, V., and Aykroyd, R.G. (2019). Birnbaum-Saunders autoregressive conditional duration models applied to high-frequency financial data. *Statistical Papers*, 60:1605–1629.
- Saulo, H., Leiva, V., Ziegelmann, F.A., and Marchant, C. (2013). A nonparametric method for estimating asymmetric densities based on skewed Birnbaum-Saunders distributions applied to environmental data. *Stochastic Environmental Research and Risk Assessment*, 27:1479–1491.
- Shumway, R.H. and Stoffer, D.S. (2017). *Time Series Analysis and its Applications with R Examples*. Springer, New York, US.
- Sim, C.H. (1987). A mixed gamma ARMA(1, 1) model for river flow time series. *Water Resources Research*, 23:32–36.
- Vila, R., Leão, J., Saulo, H., Shahzad, M.N., and Santos-Neto, M. (2020). On a bimodal Birnbaum-Saunders distribution with applications to lifetime data. *Brazilian Journal of Probability and Statistics*, pages in press.
- Villegas, C., Paula, G.A., and Leiva, V. (2011). Birnbaum-Saunders mixed models for censored reliability data analysis. *IEEE Transactions on Reliability*, 60:748–758.
- Xu, X. (2020). Forecasting air pollution PM_{2.5} in Beijing using weather data and multiple kernel learning. *Journal of Forecasting*, 39:117–125.

Appendix A: Proofs

Proof of Theorem 1 Since $\Theta(B)\Phi(B)^{-1} = \sum_{i=0}^{\infty} \psi_i B^i$ with $\psi_0 = 1$, by using (5), the BISARMA model can be rewritten as

$$\psi_t = Y_t - \mathbf{x}_t^\top \boldsymbol{\beta} = \eta + \Theta(B)\Phi(B)^{-1} \varepsilon_t. \quad (14)$$

Seeing that $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$, a.s., for all t , it follows that $E[\varepsilon_t] = 0$. Then

$$E[Y_t] = \mathbf{x}_t^\top \boldsymbol{\beta} + E[\psi_t] \stackrel{(14)}{=} \eta + \mathbf{x}_t^\top \boldsymbol{\beta} + \Theta(B)\Phi(B)^{-1}E[\varepsilon_t] = \eta + \mathbf{x}_t^\top \boldsymbol{\beta},$$

whenever the series $\Theta(B)\Phi(B)^{-1}\varepsilon_t$ exists. In addition, since $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$, a.s., for all t , and $\text{Cov}[\varepsilon_s, \varepsilon_t] = 0$ for all $t \neq s$, by (14) we have

$$\text{Var}[Y_t] = \text{Var}[\Theta(B)\Phi(B)^{-1}\varepsilon_t] = \sum_{i=0}^{\infty} \psi_i^2 \text{Var}[\varepsilon_{t-i}] = \sum_{i=0}^{\infty} \psi_i^2 E[\text{Var}[Y_{t-i} | \mathcal{F}_{t-i-1}]],$$

where in the last equality we use the law of total variance. Analogously, the covariance $\text{Cov}[Y_t, Y_{t-k}]$ can be written as $\sum_{i=0}^{\infty} \psi_i \psi_{i-k} E[\text{Var}[Y_{t-i} | \mathcal{F}_{t-i-1}]]$, $k > 0$. Thus, the proof is complete. ■

Proof of Theorem 2 By using (10), and by combining (17) and (18) (see Appendix B),

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_v} &= \sum_{t=m+1}^n \left(x_{t,v} - \sum_{i=1}^p \phi_i x_{(t-i)v} \right) \tanh\left(\frac{y_t - \mu_t}{2}\right) \left[\frac{2}{\alpha^2} \cosh^2\left(\frac{y_t - \mu_t}{2}\right) - \frac{1}{2} \right]; \\ \frac{\partial^2 \ell}{\partial \beta_v^2} &= \frac{1}{\alpha^2} \sum_{t=m+1}^n \left(x_{t,v} - \sum_{i=1}^p \phi_i x_{(t-i)v} \right)^2 \left[\frac{\alpha^2}{4} \text{sech}^2\left(\frac{y_t - \mu_t}{2}\right) - 2 \sinh^2\left(\frac{y_t - \mu_t}{2}\right) - 1 \right]. \end{aligned}$$

Note that the condition $x_{t,v} > p \max_{i=1, \dots, p} |\phi_i x_{(t-i)v}|$ implies that $x_{t,v} - \sum_{i=1}^p |\phi_i x_{(t-i)v}| > 0$. Then

$$x_{t,v} - \sum_{i=1}^p \phi_i x_{(t-i)v} > 0, \quad t = m+1, \dots, n. \quad (15)$$

Let $g(x) = \tanh(x) \left[\frac{2}{\alpha^2} \cosh^2(x) - \frac{1}{2} \right]$ and $h(x) = \frac{\alpha^2}{4} \text{sech}^2(x) - 2 \sinh^2(x) - 1$. A straightforward computation shows that $g(x)$ is an increasing function such that $g(0) = 0$ and $g(x) \rightarrow \pm\infty$ as $x \rightarrow \pm\infty$. Seeing that $x_{t,v} > p \max_{i=1, \dots, p} |\phi_i x_{(t-i)v}|$, we have that $x_{t,v} > 0$. Therefore $x_{t,v} \beta_v \rightarrow \pm\infty$ as $\beta_v \rightarrow \pm\infty$. It follows that $g\left(\frac{y_t - \mu_t}{2}\right) \rightarrow \pm\infty$ as $\beta_v \rightarrow \mp\infty$, for each $t = m+1, \dots, n$. Then, by using (15), we have

$$\frac{\partial \ell}{\partial \beta_v} \rightarrow +\infty \quad \text{as } \beta_v \rightarrow -\infty, \quad \text{and} \quad \frac{\partial \ell}{\partial \beta_v} \rightarrow -\infty \quad \text{as } \beta_v \rightarrow +\infty. \quad (16)$$

In addition, note that for $0 < \alpha < 2$, $h'(x) = -2 \frac{\sinh(x)}{\cosh^3(x)} \left(\frac{\alpha^2}{4} - 2 \cosh^4(x) \right) = 0$ iff $x = 0$, and $h''(0) = -2 \left(\frac{\alpha^2}{4} + 2 \right) < 0$. That is, $h(x)$ increases on $(-\infty, 0)$ and then decreases on $(0, \infty)$ with $h(0) = \frac{\alpha^2}{4} - 1 < 0$. Then $h(x)$ is a negative function for all $x \in \mathbb{R}$ and $0 < \alpha < 2$. Therefore, by using (15), we have that $\frac{\partial^2 \ell}{\partial \beta_v^2} < 0$, which means that $\frac{\partial \ell}{\partial \beta_v}$ is a decreasing function in β_v , for each $v = 1, \dots, k$ fixed. Finally, by using (16), the proof follows by the standard intermediate value theorem, see Lang (1998). ■

Appendix B: Fisher observed information matrix

This appendix contains expressions for the second order partial derivatives of the log-likelihood function of the BISARMA model given in (9), which can be used to obtain the elements of the Fisher observed information matrix. The first and second order derivatives of the log-likelihood function $\ell_t(\alpha, \boldsymbol{\beta}, \eta, \boldsymbol{\phi}, \boldsymbol{\theta})$ with respect to μ_t

are expressed, respectively, by

$$\begin{aligned}\frac{\partial \ell_t}{\partial \mu_t} &= \tanh\left(\frac{y_t - \mu_t}{2}\right) \left[\frac{2}{\alpha^2} \cosh^2\left(\frac{y_t - \mu_t}{2}\right) - \frac{1}{2} \right], \\ \frac{\partial^2 \ell_t}{\partial \mu_t^2} &= \frac{\partial}{\partial \mu_t} \left(\frac{\partial \ell_t}{\partial \mu_t} \right) = \frac{1}{\alpha^2} \left[\frac{\alpha^2}{4} \operatorname{sech}^2\left(\frac{y_t - \mu_t}{2}\right) - 2 \sinh^2\left(\frac{y_t - \mu_t}{2}\right) - 1 \right].\end{aligned}\quad (17)$$

Note that

$$\frac{\partial^2 \ell_t}{\partial \eta^2} = \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial \mu_t^2} = \sum_{t=m+1}^n \frac{1}{\alpha^2} \left[\frac{\alpha^2}{4} \operatorname{sech}^2\left(\frac{y_t - \mu_t}{2}\right) - 2 \sinh^2\left(\frac{y_t - \mu_t}{2}\right) - 1 \right].$$

For $\beta_{v,w}$, with $v, w = 1, \dots, k$, we have

$$\begin{aligned}\frac{\partial^2 \ell_t}{\partial \beta_v \partial \beta_w} &= \sum_{t=m+1}^n \frac{\partial}{\partial \beta_v} \left(\frac{\partial \ell_t}{\partial \beta_w} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \beta_v} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \beta_w} \right) \\ &= \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial \mu_t^2} \left(x_{tv} - \sum_{i=1}^p \phi_i x_{(t-i)v} \right) \left(x_{tw} - \sum_{i=1}^p \phi_i x_{(t-i)w} \right),\end{aligned}\quad (18)$$

where

$$\frac{\partial \mu_t}{\partial \beta_v} = x_{tv} - \sum_{i=1}^p \phi_i x_{(t-i)v}, \quad \frac{\partial \mu_t}{\partial \beta_w} = x_{tw} - \sum_{i=1}^p \phi_i x_{(t-i)w}, \quad \frac{\partial^2 \mu_t}{\partial \beta_v \partial \beta_w} = \frac{\partial}{\partial \beta_v} \left(\frac{\partial \mu_t}{\partial \beta_w} \right) = 0.$$

For β_v , with $v = 1, \dots, k$, we get

$$\begin{aligned}\frac{\partial^2 \ell_t}{\partial \beta_v \partial \eta} &= \sum_{t=m+1}^n \frac{\partial}{\partial \eta} \left(\frac{\partial \ell_t}{\partial \beta_v} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \eta} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \beta_v} \right) \\ &= \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \eta} \frac{\partial \mu_t}{\partial \beta_v} + \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial^2 \mu_t}{\partial \eta \partial \beta_v} = \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial \mu_t^2} \left(x_{tv} - \sum_{i=1}^p \phi_i x_{(t-i)v} \right).\end{aligned}$$

where $\partial^2 \mu_t / \partial \eta \partial \beta_v = \partial(\partial \mu_t / \partial \beta_v) / \partial \eta = 0$.

Notice that

$$\frac{\partial^2 \ell}{\partial \alpha^2} = -\frac{1}{\alpha^2} \sum_{t=m+1}^n \left[\frac{12}{\alpha^2} \sinh^2\left(\frac{y_t - \mu_t}{2}\right) - 1 \right].$$

For β_v , with $v = 1, \dots, k$, we reach

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \alpha \partial \beta_v} &= \sum_{t=m+1}^n \frac{\partial}{\partial \alpha} \left(\frac{\partial \ell_t}{\partial \beta_v} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \alpha} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \beta_v} \right) \\ &= -\frac{4}{\alpha^3} \sum_{t=m+1}^n \cosh\left(\frac{y_t - \mu_t}{2}\right) \sinh\left(\frac{y_t - \mu_t}{2}\right) \left(x_{tv} - \sum_{i=1}^p \phi_i x_{(t-i)v} \right).\end{aligned}$$

Notice that

$$\frac{\partial^2 \ell}{\partial \alpha \partial \eta} = \frac{\partial}{\partial \alpha} \left(\frac{\partial \ell}{\partial \eta} \right) = -\frac{4}{\alpha^3} \sum_{t=m+1}^n \sinh\left(\frac{y_t - \mu_t}{2}\right) \cosh\left(\frac{y_t - \mu_t}{2}\right).$$

For ϕ_i, ϕ_j , with $i, j = 1, \dots, k$, we have

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \phi_i \partial \phi_j} &= \sum_{t=m+1}^n \frac{\partial}{\partial \phi_i} \left(\frac{\partial \ell_t}{\partial \phi_j} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \phi_i} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \phi_j} \right) \\ &= \sum_{t=m+1}^n \left(\frac{\partial^2 \ell_t}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \phi_i} \frac{\partial \mu_t}{\partial \phi_j} + \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial^2 \mu_t}{\partial \phi_i \partial \phi_j} \right) = \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial \mu_t^2} (y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}) (y_{t-j} - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}). \end{aligned}$$

where $\partial \mu_t / \partial \phi_i = y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}$, $\partial \mu_t / \partial \phi_j = y_{t-j} - \mathbf{x}_{t-j}^\top \boldsymbol{\beta}$, $\partial^2 \mu_t / \partial \phi_i \partial \phi_j = \partial(\partial \mu_t / \partial \phi_j) / \partial \phi_i = 0$.

For ϕ_i , with $j = 1, \dots, k$, we get

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \alpha \partial \phi_i} &= \sum_{t=m+1}^n \frac{\partial}{\partial \alpha} \left(\frac{\partial \ell_t}{\partial \phi_i} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \alpha} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \phi_i} \right) \\ &= -\frac{4}{\alpha^3} \sum_{t=m+1}^n \cosh \left(\frac{y_t - \mu_t}{2} \right) \sinh \left(\frac{y_t - \mu_t}{2} \right) (y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}). \end{aligned}$$

For ϕ_i with $i = 1, \dots, k$, we reach

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \beta_v \partial \phi_i} &= \sum_{t=m+1}^n \frac{\partial}{\partial \beta_v} \left(\frac{\partial \ell_t}{\partial \phi_i} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \beta_v} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \phi_i} \right) \\ &= \sum_{t=m+1}^n \left[\frac{\partial^2 \ell_t}{\partial \mu_t^2} \left(x_{tv} - \sum_{i=1}^p \phi_i x_{(t-i)v} \right) (y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}) - \frac{\partial \ell_t}{\partial \mu_t} x_{(t-i)v} \right], \end{aligned}$$

where $\partial^2 \mu_t / \partial \beta_v \partial \phi_i = \partial(\partial \mu_t / \partial \phi_i) / \partial \beta_v = -x_{(t-i)v}$.

For ϕ_i , with $i = 1, \dots, k$, we have

$$\begin{aligned} \frac{\partial^2 \ell_t}{\partial \phi_i \partial \eta} &= \sum_{t=m+1}^n \frac{\partial}{\partial \eta} \left(\frac{\partial \ell_t}{\partial \phi_i} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \eta} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \phi_i} \right) \\ &= \sum_{t=m+1}^n \left(\frac{\partial^2 \ell_t}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \eta} \frac{\partial \mu_t}{\partial \phi_i} + \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial^2 \mu_t}{\partial \eta \partial \phi_i} \right) = \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial \mu_t^2} (y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}). \end{aligned}$$

where $\partial^2 \mu_t / \partial \eta \partial \phi_i = \partial(\partial \mu_t / \partial \phi_i) / \partial \eta = 0$.

For θ_i, θ_j , with $i, j = 1, \dots, k$, we get

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} &= \sum_{t=m+1}^n \frac{\partial}{\partial \theta_i} \left(\frac{\partial \ell_t}{\partial \theta_j} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \theta_i} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \theta_j} \right) \\ &= \sum_{t=m+1}^n \left(\frac{\partial^2 \ell_t}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \theta_i} \frac{\partial \mu_t}{\partial \theta_j} + \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial^2 \mu_t}{\partial \theta_i \partial \theta_j} \right) = \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial \mu_t^2} u_{t-i} u_{t-j}, \end{aligned}$$

where $\partial \mu_t / \partial \theta_i = u_{t-i}$, $\partial \mu_t / \partial \theta_j = u_{t-j}$, $\partial^2 \mu_t / \partial \theta_i \partial \theta_j = \partial / \partial \theta_i (\partial \mu_t / \partial \theta_j) = 0$.

For ϕ_i, θ_j , with $i, j = 1, \dots, k$, we reach

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \phi_i \partial \theta_j} &= \sum_{t=m+1}^n \frac{\partial}{\partial \phi_i} \left(\frac{\partial \ell_t}{\partial \theta_j} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \phi_i} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \theta_j} \right) \\ &= \sum_{t=m+1}^n \left(\frac{\partial^2 \ell_t}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \phi_i} \frac{\partial \mu_t}{\partial \theta_j} + \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial^2 \mu_t}{\partial \phi_i \partial \theta_j} \right) = \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial^2 \mu_t} \left(-\mathbf{x}_{t-i}^\top \boldsymbol{\beta} \right) u_{t-j}. \end{aligned}$$

where $\partial^2 \mu_t / \partial \phi_i \partial \theta_j = \partial(\partial \mu_t / \partial \theta_j) / \partial \phi_i = 0$.

For θ_j , with $j = 1, \dots, k$, we have

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \alpha \partial \theta_j} &= \sum_{t=m+1}^n \frac{\partial}{\partial \alpha} \left(\frac{\partial \ell_t}{\partial \theta_j} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \alpha} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \theta_j} \right) \\ &= -\frac{4}{\alpha^3} \sum_{t=m+1}^n \cosh \left(\frac{y_t - \mu_t}{2} \right) \sinh \left(\frac{y_t - \mu_t}{2} \right) u_{t-j} \end{aligned}$$

For θ_j , with $j = 1, \dots, k$, we get

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \beta_v \partial \theta_j} &= \sum_{t=m+1}^n \frac{\partial}{\partial \beta_v} \left(\frac{\partial \ell_t}{\partial \theta_j} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \beta_v} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \theta_j} \right) \\ &= \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \beta_v} \frac{\partial \mu_t}{\partial \theta_j} + \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial^2 \mu_t}{\partial \beta_v \partial \theta_j} = \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial^2 \mu_t} \left(x_{tv} - \sum_{i=1}^p \theta_j x_{(t-i)v} \right) u_{t-j}, \end{aligned}$$

For θ_j , with $i = 1, \dots, k$, we reach

$$\begin{aligned} \frac{\partial^2 \ell_t}{\partial \theta_j \partial \eta} &= \sum_{t=m+1}^n \frac{\partial}{\partial \eta} \left(\frac{\partial \ell_t}{\partial \theta_j} \right) = \sum_{t=m+1}^n \frac{\partial}{\partial \eta} \left(\frac{\partial \ell_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \theta_j} \right) \\ &= \sum_{t=m+1}^n \left(\frac{\partial^2 \ell_t}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \eta} \frac{\partial \mu_t}{\partial \theta_j} + \frac{\partial \ell_t}{\partial \mu_t} \frac{\partial^2 \mu_t}{\partial \eta \partial \theta_j} \right) = \sum_{t=m+1}^n \frac{\partial^2 \ell_t}{\partial \mu_t^2} u_{t-j}, \end{aligned}$$

where $\partial^2 \mu_t / \partial \eta \partial \theta_j = \partial(\partial \mu_t / \partial \theta_j) / \partial \eta = 0$.