"Bayesian analysis of linear regression models with autoregressive symmetrical errors and incomplete data"

Francyelle de Lima Medina(*)

francy@de.ufpe.br
Department of Statistics
Federal University of Pernambuco, Brazil

September, 2024
ECODEP

(*)Joint work:Aldo M. Garay, Victor H. Lachos and Suelem T. Freitas

Summary

Motivation

- A great challenge that frequently arises in practice when analyzing time series data is to know how to proceed when the series has incomplete information, such as when the observed values are not completely available.

- In the context of linear regression models with autoregressive errors and incomplete data, a routine assumption is to consider that innovations follow a normal distribution.

Motivation

- However, it is well known that the normality assumption is not always appropriate, especially when the data have outliers or heavy tails.

- An alternative is to consider a more flexible class of distributions for the error term, such as the Student-t one. As for example Lin and Lee (2007) and more recently, Liu, Kumar e Palomar (2019), Garay et al. (2020), among others.

- However, there is a lack of studies which take into account, simultaneously, incomplete information (censoring and/or missing values) and innovations modeled by a flexible class of distributions, such as, the class of scale mixture of normal (SMN) distributions.

Introduction

Main objective

We propose the linear regression model with autoregressive errors of order p and incomplete information, based on the SMN class of distributions, denoted by SMN-ILR-AR(p) model.

Source of incomplete data

We are interested in two situations:

Censored cases

Occurs when some information of the response are not completely available, but the regressive variables for these observations are completely known.

Missing cases

In this situation, we the realization y_t may be missing, i.e., $y_t = NA$ (not available).

Scale Mixtures of Normal distributions (SMN)

$$\eta = \mu + U^{-1/2} Z,$$

- U and Z are independents.
- $Z \sim N(0, \sigma^2)$.
- U is a positive random variable with cdf $H(\cdot; \boldsymbol{\nu})$ (pdf $h(\cdot; \boldsymbol{\nu})$).
- ***Notation:*** $\eta \sim SMN(\mu, \sigma^2; H)$.

Examples of SMN distributions

- ***Student-t distribution:*** $\eta \sim T(\mu, \sigma^2, \nu)$, $U \sim Gamma(\nu/2, \nu/2)$.

$$f(\eta|\sigma;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi\sigma)^{1/2}} \left(1 + \frac{\eta^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}.$$

- ***Slash distribution:*** $\eta \sim SL(\mu, \sigma^2, \nu)$,$U \sim Beta(\nu, 1)$.

$$f(\eta|\sigma;\nu) = \nu^{1/2} \int_0^1 \frac{u^{\nu-1}}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{u\eta^2}{2\sigma^2}\right\} du.$$

- ***Contaminated normal distribution:*** $\eta \sim CN(\mu, \sigma^2, \nu, \gamma)$, U is a dichotomous random variable with: $\mathrm{P}(U = \delta) = \xi$ and $\mathrm{P}(U = 1) = 1 - \xi$.

$$f(\eta|\sigma;\delta,\xi) = \xi g\left(\eta|0, \delta^{-1}\sigma^2\right) + (1-\xi) g\left(\eta|0, \sigma^2\right).$$

where g denote the density of normal distribution, with δ and $\xi \in (0, 1)$.

SMN-ILR-AR(p) model

Model formulation

Consider the linear regression model with autocorrelated symmetrical errors:

$$\begin{aligned} Y_t &= \mathrm{x}_t^\top \beta + \varepsilon_t, \qquad t = 1, 2, \dots, T, \\ \varepsilon_t &= \sum_{k=1}^{p} \phi_k \varepsilon_{t-k} + \eta_t, \quad \eta_t \sim \mathrm{SMN}(0, \sigma^2, \nu). \end{aligned} \tag{1}$$

The model can be rewritten as

$$Y_t = \mathrm{x}_t^\top \beta + \sum_{k=1}^{p} \phi_k (Y_{t-k} - \mathrm{x}_{t-k}^\top \beta) + \eta_t. \tag{2}$$

SMN-ILR-AR(p) model

Incomplete data

- **Censored cases:** (Left censored)

$$Y_{\mathrm{obs}_t} = \begin{cases} \kappa_t & \text{if} \quad Y_t \leq \kappa_t, \\ Y_t & \text{if} \quad Y_t > \kappa_t, \end{cases} \tag{3}$$

- **Missing cases**

$$y_1, \ldots, y_{t_1}, \overbrace{\mathrm{NA}, \ldots, \mathrm{NA}}^{n_1}, y_{t_1+n_1+1}, \ldots, y_{t_d}, \overbrace{\mathrm{NA}, \ldots, \mathrm{NA}}^{n_d},$$
$$y_{t_d+n_d+1}, \ldots, y_{t_D}, \underbrace{\mathrm{NA}, \ldots, \mathrm{NA}}_{n_D}, y_{t_D+n_D+1}, \ldots, y_T.$$

SMN-ILR-AR(p) model

The likelihood function

Let $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\nu}\right)^\top$ the parameters of the SMN-ILR-AR(p) model and $\mathrm{X} = \left(\mathrm{X}_1^\top, \ldots, \mathrm{X}_{n+p}^\top\right)$ the covariates, with $\mathrm{y} = (y_t, p+1 \leqslant t \leqslant n)$, $\mathrm{y}^o = (y_t, t \in C^o)$, $\mathrm{y}^c = (y_t, t \in C^c)$ and $\mathrm{y}^m = (y_t, t \in C^m)$:

$$
\begin{aligned}
f\left(\mathrm{y}_{obs} | \boldsymbol{\theta}\right) &\propto \int f\left(\mathrm{y}^o, \mathrm{y}^c, \mathrm{y}^m | \boldsymbol{\theta}\right) d\mathrm{y}^m \\
&\propto \int \left\{ \prod_{t=p+1}^{n} f_{SMN}\left(y_t | \mu_t^*, \sigma^2; \boldsymbol{\nu}\right)^{1-\mathbb{I}_{\mathbf{y}^c}(y_t)} F_{SMN}\left(\frac{\kappa_t - \mu_t^*}{\sigma}\right)^{\mathbb{I}_{\mathbf{y}^c}(y_t)} \right\} d\mathrm{y}^m,
\end{aligned}
$$

$\{Y_t | y_{t-1}, \ldots, y_{t-p}, \boldsymbol{\theta}\}$ follows a SMN distribution and $\mu_t^* = \mathrm{x}_t^\top \boldsymbol{\beta} - \sum_{i=1}^{p} \phi_i \epsilon_{t-i}$.

SMN-ILR-AR(p) model

Main results

Two important results developed by Liu et al. (2019) related to the set of M missing blocks $\mathrm{y}^m = (\mathrm{y}^{(1)}, \ldots, \mathrm{y}^{(M)})$ with $\mathrm{y}^{(k)} = \{y_{s_k+1}, \ldots, Y_{s_k+m_k}\}$, in the context of the N-ILR-AR(p) model.

(1) Given $\{\mathrm{y}_{obs}, \boldsymbol{\theta}\}$, the missing blocks y^m are independent, i.e.,

$$f\left(\mathrm{y}^m|\mathrm{y}_{obs}, \boldsymbol{\theta}\right) = \prod_{k=1}^{M} f\left(\mathrm{y}^{(k)}|\mathrm{y}_{obs}, \boldsymbol{\theta}\right). \tag{4}$$

(2) The distribution of the kth missing block $\mathrm{y}^{(k)}$, given $\{\mathrm{y}_{obs}, \boldsymbol{\theta}\}$, depends on the observed values y_{obs} only through the last p data $\mathrm{y}_{(s_k-p)} = \{y_{s_k-1}, \ldots, y_{s_k-p}\}$ and the next p observation $\mathrm{y}_{(s_k+m_k+p)} = \{y_{s_k+m_k+1}, \ldots, y_{s_k+m_k+p}\}$.

Bayesian estimation

Prior specification

Let $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\nu}\right)^\top$ and suppose independent priors:

$$\pi(\boldsymbol{\theta}) = \pi\left(\boldsymbol{\beta}^\top\right)\pi\left(\sigma^2\right)\pi(\boldsymbol{\nu})\prod_{i=1}^{p}\pi(\phi_i).$$

- $\boldsymbol{\beta} \sim \mathrm{N}_l\left(\mathrm{b}_0, \mathrm{S}_\beta\right)$.
- Considering the reparameterization of $\boldsymbol{\phi}$, used by Lin and Lee (2007), i.e.: $\gamma_i \sim \mathrm{U}(-1;1)$ for $i = 1, \ldots, p$.
- For the scale parameter, we adopt the hierarchical prior $\sigma^{-2}|\eta_2 \sim \mathrm{Gamma}(\eta_1, \eta_2)$ and $\eta_2 \sim \mathrm{Gamma}(e_1, e_2)$.
- For the $\boldsymbol{\nu}$ parameter, we use the priors suggested by Celso et. al.(2012), that is:
  a) For the Student-t and slash distributions: $\nu|\lambda \sim \mathrm{Exp}(\lambda)$ and $\lambda \sim \mathrm{U}(c, d)$.
  b) For the contaminated normal distribution: $\boldsymbol{\nu} = (\delta, \xi)^\top$, such as $\delta \sim \mathrm{Beta}(\delta_0, \delta_1)$ and $\xi \sim \mathrm{Beta}(\xi_0, \xi_1)$.

MCMC Algorithm

- *Step 1.* For each one of M missing blocks $\mathrm{y}^m = \left(\mathrm{y}^{(1)}, \ldots, \mathrm{y}^{(M)}\right)$, with $\mathrm{y}^{(k)} = \left\{y_{s_k+1}, \ldots, Y_{s_k+m_k}\right\}$, we generate $\mathrm{y}^{(*k)}$, for $k = 1, 2, \ldots M$, independently of $\pi(\mathrm{y}^{(k)}|\mathrm{y}_{(s_k-p)}, \mathrm{y}_{(s_k+m_k+p)}, \mathrm{u}, \boldsymbol{\theta})$, given by:

$$\mathrm{N}_{m_k}\left(\boldsymbol{\mu}^*_{m_k}, \boldsymbol{\Sigma}^*_{m_k}\right),$$

$$\begin{aligned} \boldsymbol{\mu}^*_{m_k} &= \mu_{m_k} + \boldsymbol{\Sigma}_{m_k \times p}\left(\boldsymbol{\Sigma}_{p\times p}\right)^{-1}\left(\mathrm{y}_{(s_k+m_k+p)} - \boldsymbol{\mu}_p\right) \\ \boldsymbol{\Sigma}^*_{m_k} &= \boldsymbol{\Sigma}_{m_k\times m_k} - \boldsymbol{\Sigma}_{m_k \times p}\left(\boldsymbol{\Sigma}_{p\times p}\right)^{-1}\boldsymbol{\Sigma}_{p\times m_k}. \end{aligned}$$

- *Step 2.* For the n_c censored observations, sample $\mathrm{y}^{*c} = (y_t^{*c},\ t \in \mathbb{C}_c)$ independently from the truncated normal distribution

$$\mathrm{TN}_{\mathbb{A}}\left(\mathrm{X}_t^\top\boldsymbol{\beta} + \sum_{j=1}^{p}\phi_j\left(y_{t-j} - \mathrm{X}_{t-j}^\top\boldsymbol{\beta}\right), \frac{\sigma^2}{u_t}\right),$$

with $\mathbb{A} = (-\infty, \kappa_t]$.

After Steps 1-2, the new $\mathrm{y}^* = vec\left(\mathrm{y}^o, \mathrm{y}^{*c}, \mathrm{y}^{*m}\right)$ represents the observed values y^o and the sample generated for the n_c censored cases and $\sum_{k=1}^{M} m_k$ missing values.

MCMC Algorithm

- *Step 3.* Sample u_t, for $t = p+1, \ldots, n$ from $\pi(u_t | \mathrm{y}^*_{(t-p)}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \mathfrak{y})$:
  1. for the Student-t case,
  $$\mathrm{Gamma}\left(\frac{\mathfrak{y}+1}{2}, \frac{(\mathrm{y}^*_t - \mu^*_t)^2}{2\sigma^2} + \frac{\mathfrak{y}}{2}\right);$$
  2. for the slash case,
  $$\mathrm{TGamma}_{\lfloor 0,1 \rfloor}\left(\mathfrak{y} + \frac{1}{2}, \frac{(\mathrm{y}^*_t - \mu^*_t)^2}{2\sigma^2} + \frac{\mathfrak{y}}{2}\right).$$
  3. for the contaminated normal case, a discrete distribution taking values δ and 1 with probabilities $\frac{p_1^*}{p_1^*+p_2^*}$ and $\frac{p_2^*}{p_1^*+p_2^*}$ respectively.
- *Step 4.* Sample $\boldsymbol{\beta}$ from $\pi(\boldsymbol{\beta} | \mathrm{y}^*, \mathrm{u}, \sigma^2, \boldsymbol{\phi}, \mathfrak{y})$, given by:
$$\mathrm{N}_l\left(\mathrm{A}_\beta^{-1}\mathrm{B}_\beta, \mathrm{A}_\beta^{-1}\right),$$

$$\mathrm{A}_\beta = \sum_{t=p+1}^{n} \frac{u_t}{\sigma^2}\left(\mathrm{X}_t^\top - \sum_{k=1}^{p} \phi_k \mathrm{X}_{t-k}^\top\right)^2 + \mathrm{S}_\beta^{-1} \text{ and } \mathrm{B}_\beta = \sum_{t=p+1}^{n} \frac{Z_t u_t}{\sigma^2}\left(\mathrm{X}_t^\top - \sum_{k=1}^{p} \phi_k \mathrm{X}_{t-k}^\top\right) + \mathrm{S}_\beta^{-1}\mathrm{b}_0,$$

MCMC Algorithm

- *Step 5.* For the scale parameter, we use:
  - (i) Sample η_2 from $\pi(\eta_2|\sigma^{-2})$, which is $\text{Gamma}\left(\eta_1 + e_1,\ \frac{1}{\sigma^2} + e_2\right)$.
  - (ii) Sample σ^{-2} from $\pi(\sigma^{-2}|\mathrm{y}^*, \mathrm{u}, \boldsymbol{\beta}, \boldsymbol{\phi}, \mathfrak{y}, \eta_2)$, which is

$$\text{Gamma}\left(\frac{n-p}{2} + \eta_1, \sum_{t=p+1}^{n} \frac{u_t}{2}\left(y_t^* - \mu_t^*\right)^2 + \eta_2\right).$$

- *Step 6.* For ϕ we use the same strategy suggested by Lin and Lee (2007)
  - (a) transform $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^\top$ to $\boldsymbol{\gamma}^* = \left(\gamma_1^*, \ldots, \gamma_p^*\right)^\top \in \mathbb{R}^p$, with

$$\gamma_i^* = \log\left\{\frac{1+\gamma_i}{1-\gamma_i}\right\}, \quad \text{for } i = 1, \ldots, p.$$

  - (b) using a M-H path, sample $\boldsymbol{\gamma}^*$ from the marginal conditional distribution.
  - (c) Having obtained $\boldsymbol{\gamma}^*$ from the M-H algorithm, we transform it back to $\boldsymbol{\gamma}$ by:

$$\gamma_i = \frac{\exp\left(\gamma_i^*\right) - 1}{\exp\left(\gamma_i^*\right) + 1}, \quad \text{for } i = 1, \ldots, p$$

and then transform $\boldsymbol{\gamma}$ back to ϕ.

Simulation studies

- We study the performance of the Bayesian estimations for the SMN-ILR-AR(p) model, considering: (i) different levels (%) of censoring, (ii) the presence of blocks with missing values and (iii) the presence of censoring and missing values simultaneously, for the response variable.

- We generated $R = 500$ artificial samples of size n, from each fixed SMN-ILR-AR(p) model, considering $p \in \{1, 2\}$.

- The real parameter values were: $\beta = (1,\ 4)^\top$, $\phi = (\phi_1, \phi_2)^\top = (0.656, -0.207)^\top$, $\sigma^2 = 1.5$ and $\mathrm{x}_t^\top = (1, x_t)^\top$ with x_t, $t = 1, 2, \ldots, n$ as a random sample from a uniform distribution in the interval $(1,\ 3)$.

- We fitted the model using the proposed MCMC algorithm and obtaining $100,000$ MCMC draws from the joint posterior distribution of parameters, where the first $20,000$ (20%) iterations were discarded as burn-in samples.

Simulation studies

- We analyzed the parameter estimation of the SMN-ILR-AR(2) model, with $n = 300$, under the presence of $B \in \{1, 3\}$ blocks with missing values and different censoring proportions (5% and 15%) simultaneously.

- We consider the average values (**MC-Mean**) and standard deviations of the estimates (**MC-Sd**), computed across all R samples.

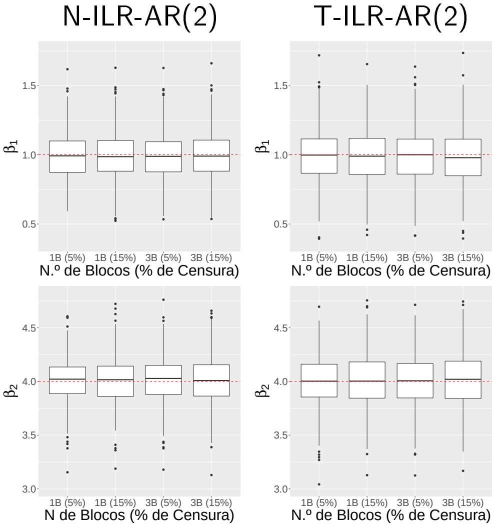
- **MC-Cov** is the percentage of times, computed across all samples, that the 95% highest posterior density intervals contain the true value of the parameter.

Simulation studies

# of Blocks NA	Cens. Level		N-ILR-AR(2) MC-Mean	N-ILR-AR(2) MC-Sd	N-ILR-AR(2) MC-Cov	T-ILR-AR(2) MC-Mean	T-ILR-AR(2) MC-Sd	T-ILR-AR(2) MC-Cov
1 block (1.6%)	5%	β_1	0.9926	0.1678	96.2%	0.9940	0.1962	95.6%
		β_2	4.0154	0.2041	95.8%	4.0018	0.2392	95.4%
		ϕ_1	0.6412	0.0592	94.2%	0.6391	0.0540	95.0%
		ϕ_2	-0.1962	0.0593	93.6%	-0.1947	0.0518	95.0%
		σ^2	1.5476	0.1366	94.2%	1.5054	0.2184	93.6%
	15%	β_1	0.9948	0.1783	95.6%	0.9880	0.2076	95.4%
		β_2	4.0130	0.2248	95.8%	4.0121	0.2589	97.0%
		ϕ_1	0.6106	0.0630	92.0%	0.6150	0.0578	92.4%
		ϕ_2	-0.1756	0.0630	94.2%	-0.1740	0.0546	91.4%
		σ^2	1.5966	0.1567	92.0%	1.5856	0.2434	93.2%
3 blocks (5%)	5%	β_1	0.9906	0.1708	95.6%	0.9933	0.1996	94.8%
		β_2	4.0180	0.2081	94.0%	4.0048	0.2445	96.2%
		ϕ_1	0.6424	0.0603	94.2%	0.6371	0.0548	94.2%
		ϕ_2	-0.1969	0.0606	94.4%	-0.1933	0.0531	94.4%
		σ^2	1.5478	0.1392	93.6%	1.5039	0.2228	94.0%
	15%	β_1	0.9950	0.1818	96.0%	0.9842	0.2116	95.8%
		β_2	4.0164	0.2304	95.4%	4.0184	0.2653	97.2%
		ϕ_1	0.6069	0.0642	90.2%	0.6094	0.0593	92.0%
		ϕ_2	-0.1734	0.0644	93.6%	-0.1711	0.0559	92.4%
		σ^2	1.6015	0.1604	91.2%	1.5868	0.2484	94.8%

$\beta = (\beta_1, \beta_2)^\top = (1,\ 4)^\top$, $\phi = (\phi_1, \phi_2)^\top = (0.656, -0.207)^\top$ e $\sigma^2 = 1.5$.

Simulation studies



Application

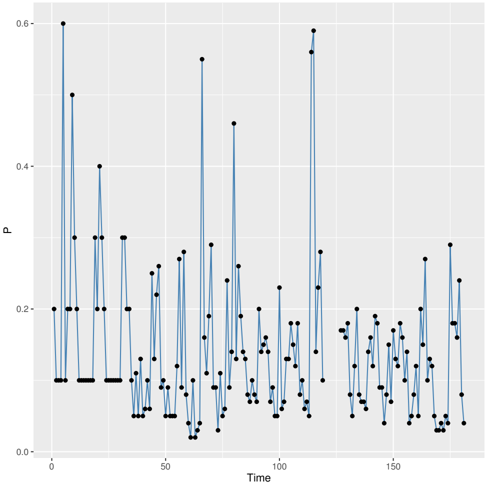
Data Description:

- We used the Phosphorus Concentration Data of West Fork Cedar River in Finchford, Iowa (USA), previously analyzed by Wang e Chan (2018) and Schumacher, Lachos e Dey (2017).

- This dataset consists of $n = 181$ observations of phosphorus concentration (P) in mg/l, and the water discharge (Q) obtained monthly from October/1998 to October/2013.

- The data presents 28 left censored observations and 7 missing values (NA), which represent approximately 15.5% and 3.9% of the total, respectively. The gap from 09/2008 to 03/2009 in the data was due to program suspension owing to lack of funding.

Application



Application

- Considering the presence of seasonality in the data, as suggested by Wang and Chan (2018) and Schumacher et al. (2017), we consider the following model:

$$\log(\mathrm{P}_t) = \sum_{j=1}^{4}[\beta_{1,j}S_{j,t} + \beta_{2,j}S_{j,t}\log(\mathrm{Q}_t)] + \varepsilon_t$$

where $\mathrm{S}_t = (S_{t1}, S_{t2}, S_{t3}, S_{t4})$ is a quarter indicator variable, such that:

$$S_{tj} = \begin{cases} 1, & \text{if the t-th observation belongs} \\ & \text{to the j-th quarter.} \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

for $t = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, 4$;

Application

- We decided to consider in our analysis each SMN-ILR-AR(p) model with $p \in \{1, 2, 3\}$, for comparative purposes.

- For each SMN-ILR-AR(p) model, we generated $150,000$ draws from the posterior distribution, discarded the first $30,000$ iterations as a burn-in and considered spacing of size 5 to reduce autocorrelation.

- We used Bayes factors to compare the models, involving the calculation of marginal densities of the data, $f(\mathrm{y}|\theta_k)$, in each SMN-ILR-AR(p) model (M_k).

Tabela: Estimated marginal densities in several SMN-ILR-AR(p) models.

Model (M_k)		$f(\mathrm{y}\|M_k)$	Model (M_k)		$f(\mathrm{y}\|M_k)$
N-ILR	AR(1)	$8.1636 \times e^{-63}$	SL-ILR	AR(1)	$3.4479 \times e^{-64}$
	AR(2)	$3.1168 \times e^{-63}$		AR(2)	$5.6672 \times e^{-62}$
	AR(3)	$9.0565 \times e^{-64}$		AR(3)	$2.1892 \times e^{-62}$
T-ILR	AR(1)	$7.3787 \times e^{-64}$	CN-ILR	AR(1)	$2.8059 \times e^{-71}$
	AR(2)	$1.5810 \times e^{-62}$		AR(2)	$3.0997 \times e^{-63}$
	AR(3)	$1.2164 \times e^{-61}$		AR(3)	$4.4727 \times e^{-69}$

Application

- The 12 latest values were preserved for prediction comparison purposes and the remaining were used for estimation and one-step-ahead prediction.

- As recommended by Lin and Lee (2007) and Garay et al. (2020), in the context of dependent longitudinal data, a more appropriate measure of fitness is the predictive accuracy of future observations. Thus, considering the best five fitted models selected by the Bayes factors.

- We used the mean absolute error (MAE) and mean square predicted error (MSPE), which are defined by:

$$
\begin{aligned}
\mathrm{MSPE}_{M_k} &= \frac{1}{12}\sum_{h=170}^{181}\left(\log(\mathrm{P}_h) - \widehat{\log(\mathrm{P}_h)}_{M_k}\right)^2 \text{ and} \\
\mathrm{MAE}_{M_k} &= \frac{1}{12}\sum_{h=170}^{181}|\log(\mathrm{P}_h) - \widehat{\log(\mathrm{P}_h)}_{M_k}|,
\end{aligned}
$$

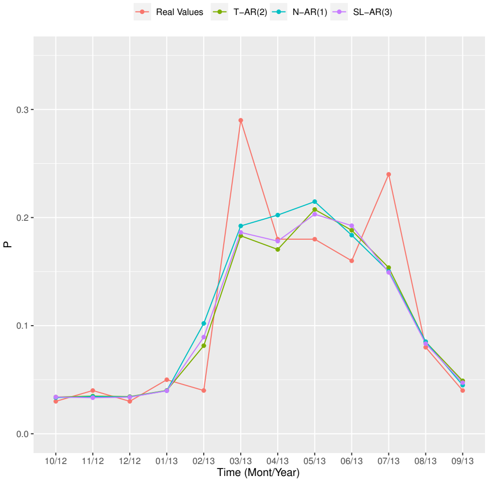
where $\widehat{\log(\mathrm{P}_h)}_{M_k}$ denotes the predicted value at the h-th, considering the M_k model.

Application

Tabela: MPSE and MA for some SMN-ILR-AR(p) models.

Model (M_k)	**MSPE**	**MAE**
N-ILR-AR(1)	0.12074	0.25326
T-ILR-AR(2)	**0.09260**	**0.23798**
T-ILR-AR(3)	0.09893	0.24155
SL-ILR-AR(2)	0.09946	0.23820
SL-ILR-AR(3)	0.10523	0.24189

Application

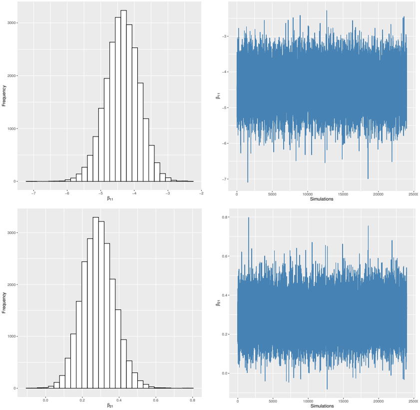


Application

Tabela: MCMC summaries for the T-ILR-AR(2) model.

θ	MCMC Summaries			
	Mean	SD	$Q_{2,5\%}$	$Q_{97,5\%}$
$\beta_{1,1}$	-4.3545	0.4979	-5.3643	-3.4032
$\beta_{1,2}$	-2.7769	0.7478	-4.3002	-1.3720
$\beta_{1,3}$	-4.2577	0.4181	-5.0944	-3.4431
$\beta_{1,4}$	-5.0052	0.4941	-5.9894	-4.0430
$\beta_{2,1}$	0.2868	0.0890	0.1182	0.4672
$\beta_{2,2}$	0.1425	0.1038	-0.0528	0.3548
$\beta_{2,3}$	0.3839	0.0690	0.2488	0.5215
$\beta_{2,4}$	0.4173	0.0918	0.2389	0.6007
ϕ_1	-0.0535	0.0820	-0.1880	0.2415
ϕ_2	0.1175	0.0745	0.0252	0.2667
σ^2	0.1562	0.0336	0.0975	0.2300
ν	3.8277	1.3646	2.0522	7.1866

Application



Conclusions

- We have introduced a robust linear regression model with AR(p) symmetrical errors, and incomplete information - SMN-IL-AR(p) - which is an extension of the models presented by Schumacher et al. (2017) and Liu et al. (2019).
- We developed an efficient MCMC algorithm for conducting Bayesian inference and its performance was evaluated through simulations and analysis of a real dataset on phosphorus concentration in river water.
- The implementation of the algorithm was done using the R package.

References

- ANDREWS, D. F.; MALLOWS, C. L. Scale mixtures of normal distributions. Journal of the Royal Statistical Society: Series B (Methodological), Wiley Online Library, v. 36, n. 1, p. 99–102, 1974.
- BARNDORFF-NIELSEN, O.; SCHOU, G. On the parametrization of autoregressive models by partial autocorrelations. Journal of multivariate Analysis, Elsevier, v. 3, n. 4, p. 408–419, 1973.
- CABRAL, C. R. B.; LACHOS, V. H.; MADRUGA, M. R. Bayesian analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population. Journal of Statistical Planning and Inference, Elsevier, v. 142, n. 1, p. 181–200, 2012.
- GAMERMAN, D.; LOPES, H. F. Markov chain Monte Carlo: stochastic simulation for Bayesian inference. Boca Raton: CRC press, 2006.
- GARAY, A. M.; BOLFARINE, H.; LACHOS, V. H.; CABRAL, C. R. Bayesian analysis of censored linear regression models with scale mixtures of normal distributions. Journal of Applied Statistics, Taylor & Francis, v. 42, n. 12, p. 2694–2714, 2015.

References

- GARAY, A. M.; MEDINA, F. L.; CABRAL, C. R.; LIN, T.-I. Bayesian analysis of the porder integer-valued ar process with zero-inflated poisson innovations. Journal of Statistical Computation and Simulation, Taylor & Francis, v. 90, n. 11, p. 1943–1964, 2020.

- LIN, T. I.; LEE, J. C. Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. Journal of Statistical Planning and Inference, Elsevier, v. 137, n. 2, p. 484–495, 2007.

- LIU, J.; KUMAR, S.; PALOMAR, D. P. Parameter estimation of heavy-tailed ar model with missing data via stochastic em. IEEE Transactions on Signal Processing, IEEE, v. 67, n. 8, p. 2159–2172, 2019.

- SCHUMACHER, F. L.; LACHOS, V. H.; DEY, D. K. Censored regression models with autoregressive errors: A likelihood-based perspective. Canadian Journal of Statistics, Wiley Online Library, v. 45, n. 4, p. 375–392, 2017.

- WANG, C.; CHAN, K.-S. Quasi-likelihood estimation of a censored autoregressive model with exogenous variables. Journal of the American Statistical Association, Taylor & Francis, v. 113, n. 523, p. 1135–1145, 2018.

Thank you.