Robust Local Bootstrap for Periodogram Statistics with Application to the Air Quality

Carlo Corrêa Solci[1], Valdério Anselmo Reisen[1,2], Paulo Canas Rodrigues[3]

[1] Universidade Federal do Espírito Santo

[2] CentraleSupeléc

[3] Universidade Federal da Bahia

Abbreviated abstract: The aim of this study is to propose a generalization of the local bootstrap for periodogram statistics to the case when causal time series are contaminated by additive outliers. In order to achieve robustness, we suggest to replace the classical version of the periodogram with the robust *M*-periodogram in the bootstrap procedure. The performance of the proposed robust methodology was compared with the classical one via a Monte Carlo experiment. The daily mean concentration of the particulate matter (PM_{10}) data in the Great Vitória Region, Brazil, was used as a real application of the approaches in the air quality area.

Related publications:

– E. Paparoditis and D. N. Politis, Journal of Time Series Analysis 20 (2), 193-222 (1999)

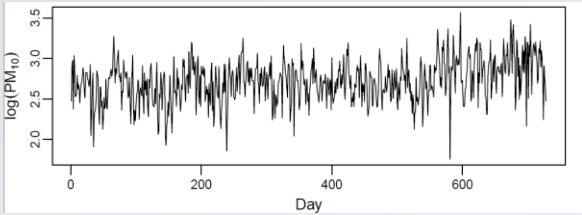
– F. A. Fajardo et al., Statistics 52 (5), 665-683 (2018)

Universidade Federal do Espírito Santo

3rd Conference on Statistics and Data Science
Salvador, Brazil (online)
October 28-30, 2021

Problem, Data, Previous Works

- **Main Motivation:** Occurrence in real applications of outilers. The Figure below display this phenomenon in the Inhalable Particulate Matter (PM_{10}) data observed in Vitória, Brazil.



- Why Robust Methods? In general, the occurrence and the position of the outliers are unknown, so that the treatment of such observations may be a difficult task. Therefore an attractive alternative is to perform robust inference which is not affected by outliers.
- Robustness for Local Bootstrap for Periodogram Statistics: Paparoditis and Politis (1999) proposed the classical version of the local bootstrap in the periodogram and Fajardo et el. (2018) developed a robust version of the periodogram, known as the *M*-periodogram. However, a robust version of the local bootstrap for periodogram statistics has not been done yet.
- Idea: We propose the use of the methodology of Paparoditis and Politis (1999) in the *M*-periodogram of Fajardo et al. (2018) as a robust method to estimate confidence intervals of time series models. The proposed robust approach is compared with the classical one through Monte Carlo experiments.

Methods

Given a sample $Z_1, Z_2, \dots, Z_N$ of a causal stochastic process $\{Z_t\}_{t\in\mathbb{Z}}$ with additive outliers, $I_{N,\psi}(\lambda_j)$ can be related to the *M*-estimator $\widehat{\boldsymbol{\beta}}_{N,\psi}(\lambda_j) = \left(\hat{\beta}_{N,\psi}^{(1)}(\lambda_j), \hat{\beta}_{N,\psi}^{(2)}(\lambda_j)\right)'$ of $\boldsymbol{\beta}_N(\lambda_j) = \left(\beta_N^{(1)}(\lambda_j), \beta_N^{(2)}(\lambda_j)\right)'$ in the linear model

$$Z_i = c'_{N,i}(\lambda_j)\boldsymbol{\beta}_N(\lambda_j) = \beta_N^{(1)}(\lambda_j)\cos(i\lambda_j) + \beta_N^{(2)}(\lambda_j)\sin(i\lambda_j) + \varepsilon_i, \quad 1 \le i \le N,$$

where $c'_{N,i}(\lambda_j) = (\cos(i\lambda_j), \sin(i\lambda_j))$. The *M*-estimator $\widehat{\boldsymbol{\beta}}_{N,\psi}(\lambda_j)$ is defined as the solution of

$$\sum_{i=1}^{n} c'_{N,i}\, \psi(Z_i - c'_{N,i}\, \widehat{\boldsymbol{\beta}}_{N,\psi}(\lambda_j)) = \mathbf{0},$$

where $\psi(\cdot)$ was chosen as the Huber function. Hence, it can be shown that the *M*-periodogram is given by

$$I_{N,\psi}(\lambda_j) = \frac{N}{8\pi}\left\|\widehat{\boldsymbol{\beta}}_{N,\psi}(\lambda_j)\right\|^2 = \frac{N}{8\pi}[(\hat{\beta}_{N,\psi}^{(1)}(\lambda_j))^2 + (\hat{\beta}_{N,\psi}^{(2)}(\lambda_j))^2].$$

The bootstrap replicates $I^*_{N,\psi}(\lambda_j)$, j $= 0,1,\dots,N'$ of the robust periodogram can be obtained via the algorithm.

1. Choose a resampling width $k_{N,\psi}$ where $k_{N,\psi} = k_\psi(N) \in \mathbb{N}$ and $k_{N,\psi} \le [N'/2]$.
2. Define i.i.d. discrete random variables $J_{1,\psi}, J_{2,\psi}, \dots, J_{N',\psi}$ that assume values in $\{-k_{N,\psi}, -k_{N,\psi}+1, \dots, k_{N,\psi}\}$ with probability $P(J_{i,\psi} = s) = p_{k_{N,\psi},s}$ for $s = 0, \pm 1, \dots, \pm k_{N,\psi}$.
3. The robust bootstrap periodogram can be defined by $I^*_{N,\psi}(\lambda_j)$= $I^*_{N,\psi}(\lambda_{J_{j,\psi}+j})$ for j $= 0,1,\dots,N'$, $I^*_{N,\psi}(\lambda_j)$= $I^*_{N,\psi}(-\lambda_j)$ for $\lambda_j < 0$ and for $\lambda_j = 0$ we have that $I^*_{N,\psi}(\lambda_j)$=0.

Assumption: $k_{N,\psi} \to \infty$ as $N \to \infty$ such that $k_{N,\psi} = o(N)$. Moreover, $\sum_{s=-k_{\psi,N}}^{k_{\psi,N}} p_{k_{\psi,N}} = 1$, $p_{k_{N,\psi},s}$= $p_{k_{N,\psi},-s}$ and $\sum_{s=-k_{\psi,N}}^{k_{\psi,N}} p_{k_{\psi,N}}^2 \to 0$ as $k_{N,\psi} \to \infty$.

Universidade Federal do Espírito Santo

3rd Conference on Statistics and Data Science
Salvador, Brazil (online)
October 28-30, 2021

Results and Conclusions

- Monte Carlo Study

Samples $Y_1, \ldots, Y_N$ ($N = 300$) of SARIMA$(1,0,0) \times (1,0,0)_S$ processes $Y_t = \phi Y_{t-1} + \Phi Y_{t-S} - \phi\Phi Y_{t-S-1} + \epsilon_t$ with $S = 4$, $\phi = 0.3$, $\Phi = 0.3$ and $\{\epsilon_t\} \sim i.i.d.\ N(0,1)$ were generated through 400 trials with 2000 bootstrap replicates each one with $P_{2k_N+1} = \{0.2, 0.2, 0.2, 0.2, 0.2\}$. Scenarios: (i) the samples are uncontaminated (ii) the samples are contaminated by $Z_t = Y_t + \omega V_t$, were $\omega = 7$ and $\{V_t\}$ are i.i.d. random variables assuming the values $-1,0,1$ with $P(V_t = 0) = 1 - \xi$, $P(V_t = -1) = P(\mathrm{V}_t = 1) = \xi/2$ where $\xi = 0.01$.

ω	I_X^*	$\overline{\hat{\phi}^*_{mean}}$	$\overline{CI(\hat{\phi}^*)}$	$PC(\hat{\phi}^*)$	$\overline{\hat{\Phi}^*_{mean}}$	$\overline{CI(\hat{\Phi}^*)}$	$PC(\hat{\Phi}^*)$
0	C	0.2811	(0.1812,0.3749)	0.9050	0.2883	(0.1896,0.3808)	0.9250
	M	0.2653	(0.1646,0.3604)	0.8700	0.2732	(0.1739,0.3668)	0.8825
7	C	0.2012	(0.1007,0.2978)	0.4825	0.1999	(0.0999,0.2954)	0.4600
	M	0.2616	(0.1610,0.3564)	0.8250	0.2571	(0.1571,0.3513)	0.8275

I_X^*	$\hat{\phi}_1$	$CI_{95\%}(\hat{\phi}_1)$	$\hat{\Phi}_1$	$CI_{95\%}(\hat{\Phi}_1)$	$\hat{\Phi}_2$	$CI_{95\%}(\hat{\Phi}_2)$
C	0.3924	(0.3176,0.4480)	0.0699	(-0.0544,0.1783)	0.1015	(-0.0279,0.2017)
M	0.3948	(0.3226,0.4480)	0.0764	(-0.0503,0.1833)	0.1228	(-0.0132,0.2193)

- Application

We have applied these methodologies to the daily average concentrations of the PM_{10} pollutant of Jardim Camburi, Vitória, Brazil from January 1st, 2018 to December 29th, 2019. Firstly, a deterministic trend was removed utilizing b-splines and then the BIC criterion was used to obtain the orders $p = 1$ and $P = 2$ of the model SARIMA$(p, 0,0) \times (P, 0,0)_S$. The bootstrap was with $P_{2k_N+1} = \{1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9\}$.

- Conclusions

All the results presented in this study give strong motivation to use the proposed robust methodology in practical situations in which causal time series contain additive outliers.

Universidade Federal do Espírito Santo

3rd Conference on Statistics and Data Science
Salvador, Brazil (online)
October 28-30, 2021