Robust Local Bootstrap for Stationary Time Series with Missing Data

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

[1] Federal University of Espírito Santo

[2] CentraleSupeléc

[3] Federal University of Bahia

Abbreviated abstract: This study proposes a generalization of the classical and of the robust versions of the local bootstrap for periodogram statistics for incomplete stationary time series by replacing the series with its amplitude modulated version. A Monte Carlo experiment was conducted to compare the performance of the proposed bootstrap methodologies, to estimate 95% confidence intervals for the parameters of autoregressive time series via the Whittle estimator. The daily mean concentration of the particulate matter (PM10) data was used to illustrate the proposed methodologies in a real application.

Related publications:

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

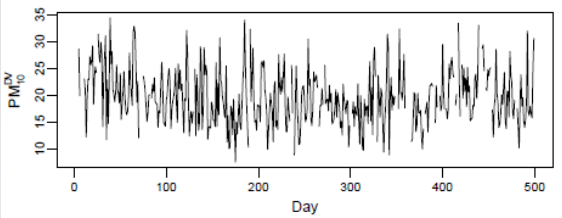
– E. Parzen, Sankhyā: The Indian Journal of Statistics, Series A 25 (4), 383-392 (1963)

Universidade Federal do Espírito Santo

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Problem, Data, Previous Works

- Main Motivation: Occurrence in real applications of missing data and outliers. The Figure below display this phenomenon in the Inhalable Particulate Matter (PM10) data observed in Vitória, Brazil.



- Why Amplitude Modulation? The presence of missing data makes a time series no longer have equally spaced observations. Therefore an attractive alternative is to replace the series by its amplitude modulated version which can be handled as a stationary sequence of equally spaced observations.
- Robustness for Local Bootstrap with Missing Data: Paparoditis and Politis (1999) proposed the local bootstrap in the periodogram, Fajardo et al. (2018) proposed the robust *M*-periodogram and Parzen (1963) developed the amplitude modulation theory. However, a classical and a robust version of the local bootstrap for time series in the presence of missing data has not been developed yet.
- Idea: We propose the use of the local bootstrap in the periodogram and in the *M*-periodogram of the amplitude modulated version of a time series as, respectively, a classical and a robust method to estimate confidence intervals of parameters of time series models in the presence of missing data. The proposed approaches are compared with each other through Monte Carlo experiments.

Methods

If only the observations at times $1 < n_1 < n_2 < \cdots < n_M < N$ of a sample $X_1, X_2, \ldots, X_N$ of a causal stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ are available, one alternative is to define the amplitude modulated sequence

$$Y_t = a_t X_t, \qquad t = 1,2,\ldots,N,$$

where

$$a_t = \begin{cases} 1, & \text{if } t = n_j \text{ for some } j \\ 0, & \text{if } t \neq n_j \text{ for all } j \end{cases}$$

The bootstrap replicates $I_N^*(\lambda_j)$, j $= 0,1,\ldots,N-1$ of the classical $I_Y(\lambda_j)$ and the robust $I_{Y,\psi}(\lambda_j)$ periodograms can be obtained by letting, respectively, $I_N(\lambda_j) = I_Y(\lambda_j)$ and $I_N(\lambda_j) = I_{Y,\psi}(\lambda_j)$ in the following algorithm:

1. Choose a resampling width k_N where $k_N = k(N) \in \mathbb{N}$ and $k_N \leq [N'/2]$ with $N' = [N/2]$.
2. Define i.i.d. discrete random variables $J_1, J_2, \ldots, J_{N'}$ that assume values in $\{-k_N, -k_N+1, \ldots, k_N\}$ with probability $P(J_i = s) = p_{k_N,s}$ for $s = 0, \pm 1, \ldots, \pm k_N$.
3. The bootstrap periodogram in the presence of missing data can be defined by $I_N^*(\lambda_j)$= $I_N\ (\lambda_{J_j+j})$ for j $=$ $0,1,\ldots,N'$, $I_N^*(\lambda_j)$= $I_N^*(2\pi - \lambda_j)$ for $N'+1 \leq j \leq N-1$ and $I_N^*(\lambda_j) = 0$ for $\lambda_j = 0$.

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

   $p_{k_N,s} = p_{k_N,-s}$ and $\sum_{s=-k_N}^{k_N} p_{k_N}^2 \to 0$ as $k_N \to \infty$.

Universidade Federal do Espírito Santo

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Results and Conclusions

- Monte Carlo Study

Samples $X_1, \dots, X_N$ ($N = 200$) of AR(1) processes $X_t = \phi X_{t-1} + \epsilon_t$ with $\phi = 0.5$ and $\{\epsilon_t\} \sim i.i.d.\ N(0,1)$ were generated via 1000 trials with 5000 bootstrap replicates. Let $Z_t = Y_t + \omega V_t$, where $Y_t = a_t X_t$ with $P(a_t = 1) = pr_{nm}$ and $P(a_t = 0) = 1 - pr_{nm}$ where $pr_{nm} = 1$ and 0.95, $\omega = 0$ and 4, while $\{V_t\}$ are i.i.d. r.v. taking the values $-1,0,1$ with $P(V_t = 0) = 1 - pr_{out}$, $P(V_t = -1) = P(\mathrm{V}_t = 1) = pr_{out}/2$ where $pr_{out} = 0$ and 0.01.

Bootstrap Estimates for $\phi = 0.5$ with $REP_{MC} = 1000$, $B = 5000$, $pr_{nm} = 1$, $pr_{out} = 0$ and $N = 200$.

ω	I_N^*	$\overline{\hat{\phi}^*}$	$\overline{\mathrm{SD}(\hat{\phi}^*)}$	$\overline{\gamma_1(\hat{\phi}^*)}$	$\overline{\mathrm{CI}_{95\%}(\hat{\phi}^*)}$	$\overline{\mathrm{A}(\hat{\phi}^*)}$	$\overline{\mathrm{P}(\hat{\phi}^*)}$
0	C	0.4754(0.0610)	0.0482(0.0086)	-0.2857(0.0921)	(0.3449,0.5882)	0.2433	0.9450
	M	0.4546(0.0636)	0.0492(0.0088)	-0.2729(0.0933)	(0.3241,0.5774)	0.2533	0.9450
	C_m	0.4779(0.0606)	0.0479(0.0086)	-0.2878(0.0968)	(0.3469,0.5901)	0.2432	0.9480
	M_m	0.4572(0.0632)	0.0490(0.0087)	-0.2703(0.0913)	(0.3266,0.5785)	0.2519	0.9450

Bootstrap Estimates for $\phi = 0.5$ with $REP_{MC} = 1000$, $B = 5000$, $pr_{nm} = 0.95$, $pr_{out} = 0.01$ and $N = 200$.

ω	I_N^*	$\overline{\hat{\phi}^*}$	$\overline{\mathrm{SD}(\hat{\phi}^*)}$	$\overline{\gamma_1(\hat{\phi}^*)}$	$\overline{\mathrm{CI}_{95\%}(\hat{\phi}^*)}$	$\overline{\mathrm{A}(\hat{\phi}^*)}$	$\overline{\mathrm{P}(\hat{\phi}^*)}$
0	C_m	0.4772(0.0641)	0.0520(0.0105)	-0.2941(0.0988)	(0.3411,0.5975)	0.2564	0.9470
	M_m	0.4546(0.0671)	0.0529(0.0106)	-0.2763(0.0981)	(0.3124,0.5751)	0.2627	0.9460
4	C_m	0.4269(0.0757)	0.0545(0.0110)	-0.2600(0.1011)	(0.2705,0.5638)	0.2933	0.8950
	M_m	0.4359(0.0690)	0.0540(0.0106)	-0.2625(0.0994)	(0.2974,0.5646)	0.2672	0.9420

- Application

We have applied the proposed methodologies to the PM_{10} of Downtown Vitória from 01/01/2018 to 05/15/2019.

Bootstrap estimates of the 95% confidence interval of the AR(1) coefficients for the $\mathrm{PM}_{10}^{\mathrm{DV}}$ time series.

I_Y^*	$\mathrm{CI}_{95\%}(\hat{\phi}^*)$
C_m	(0.2109, 0.3649)
M_m	(0.2486, 0.4043)

- Conclusions

The classical and the robust proposed methodologies are an alternative to estimate confidence intervals of parameters of stationary series, respectively, without and with additive outliers.

Universidade Federal do Espírito Santo

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022