An Overview of Robust Spectral Estimators

Valdério Reisen, Céline Lévy-Leduc, Higor Cotta, Pascal Bondon, Márton Ispány, Paulo Prezotti

An overview of robust spectral estimators

Valdério Anselmo Reisen[1,3], Céline Lévy-Leduc[2], Higor Henrique Aranda Cotta[1,3], Pascal Bondon[3], Marton Ispany[4], and Paulo Roberto Prezotti Filho[3,5]

[1] DEST and PPGEA-Universidade Federal do Espirito Santo-UFES, Vitória, Brazil
valderioanselmoreisen@gmail.com, valderio.reisen@ufes.br
[2] UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, Paris, France
[3] Laboratoire des Signaux et Systémes (L2S), CNRS-CentraleSupélec-Université Paris-Sud, Gif sur Yvette, France
[4] University of Debrecen, Debrecen, Hungary
[5] PPGEA-UFES and IFES, Brazil

Abstract. The periodogram function is widely used to estimate the spectral density of time series processes and it is well-known that this function is also very sensitive to outliers. In this context, this paper deals with robust estimation functions to estimate the spectral density of univariate and periodic time series with short and long-memory properties. The two robust periodogram functions discussed and compared here were previously explicitly and analytically derived in Fajardo et al. (2018), Reisen et al. (2017) and Fajardo et al. (2009) in the case of long-memory processes. The first two references introduce the robust periodogram based on $M-$regression estimator. The third reference is based on the robust autocovariance function introduced in Ma and Genton (2000) and studied theoretically and empirically in Lévy-Leduc et al. (2011). Here, the theoretical results of these estimators are discussed in the case of short and long-memory univariate time series and periodic processes. A special attention is given to the $M-$ periodogram for short-memory processes. In this case, Theorem 1 and Corollary 1 derive the asymptotic distribution of this spectral estimator. As the application of the methodologies, robust estimators for the parameters of AR, ARFIMA and PARMA processes are discussed. Their finite sample size properties are addressed and compared in the context of absence and presence of atypical observations. Therefore, the contributions of this paper come to fill some gaps in the literature of modeling univariate and periodic time series to handle additive outliers.

Time series, M-estimation, Q_N-estimation, long-memory, periodic processes, robustness.

1 Introduction

It is well known that outlying observations may completely destroy most of the standard estimators and several authors developed robust approaches in order to mitigate the impact of additive outliers, specially in time series models which is the process considered in this paper. However, most of the work is devoted to the robust estimation of the location, scale and other statistical tools. In this direction, the classical periodogram is the natural estimator of the spectral density of a time series and recent studies indicate that the periodogram is highly sensitive to the presence of outliers, and, thus, it becomes useless in any sub-sequential analysis. As a viable approach to attenuate this issue, the $M-$regression method applied to build alternative spectral estimators given in Fajardo et al. (2018) and Reisen et al. (2017)) and the Q_N-periodogram introduced in Fajardo et al. (2009) are some methodologies proposed recently in the literature of time series to handle additive outliers.

The M-periodogram is discussed in Fajardo et al. (2018) and Reisen et al. (2017) for the long-memory time series. The short-range process was still an open problem and is one main contribution of this paper. The asymptotic property of the $M-$periodogram is derived for the process which is identified to have short-memory property such as an ARMA model (Theorem 1). As a second contribution of this paper, the recent results given Fajardo et al. (2018) and Reisen et al. (2017), for long-memory model, are summarized and these methods are compared empirically with Q_N-periodogram and the classical periodogram which is widely used in modelling time series data. Here, these methods are empirically studied and compared in time series with and without additive outliers with the aim to verify their finite sample size robustness properties, that is, to verify their capacity to accommodate the additive outlier's effect.

The use of $M-$ and Q_N-periodograms in periodic ARMA (PARMA) models is also discussed here in the context of handling atypical or aberrant observations (additive outliers). This becomes the third contribution of this paper.

This paper is organized as follows: Section 2 discusses robust periodograms based on $M-$regression method and Q_N function for short and long-memory time series. Section 3 presents some simulation results for the methods discussed in Section 2. Section 4 gives some applications of the alternative periodograms in short and long-memory and periodic processes.

2 Robust periodograms

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a second order stationary process. Since this paper deals with short and long-memory processes, additional assumptions on the process $\{Y_t\}_{t\in\mathbb{Z}}$ will be given in the sequel of the paper. For a sample $\{Y_1, Y_2, ..., Y_N\}$, the classical periodogram function, at the Fourier frequency $\lambda_j = 2\pi j/N, j = 1, \ldots, [N/2]$, is defined as

$$I_N(\lambda_j) = \frac{1}{2\pi N}\left|\sum_{k=1}^{N} Y_k \exp(\mathrm{i}k\lambda_j)\right|^2 . \tag{1}$$

Next subsections deal with alternative periodogram functions which present similar performance (from theoretical and empirical meaning) to $I_N(\lambda)$, $\lambda \in (-\pi, \pi)$, but with robustness property against additive outliers and asymmetric and heavy-tail distributions.

2.1 *M*-periodogram

One alternative way to derive the periodogram function $I_N(\lambda_j)$ is based on the Least Square (LS) estimates of a two-dimensional vector $\boldsymbol{\beta'} = (\beta^{(1)}, \beta^{(2)})$ in the linear regression model

$$Y_i = c'_{Ni}\boldsymbol{\beta} + \varepsilon_i = \beta^{(1)}\cos(i\lambda_j) + \beta^{(2)}\sin(i\lambda_j) + \varepsilon_i \ , \ 1 \le i \le N, \ \boldsymbol{\beta} \in \mathbb{R}^2 \ , \tag{2}$$

where ε_i denotes the deviation of Y_i from $c'_{Ni}\boldsymbol{\beta}$ and $\mathbb{E}(\varepsilon_i) = 0$ and $\mathbb{E}(\varepsilon_i^2) < \infty$. In the sequel (ε_i) is assumed to be a function of a stationary Gaussian process, see (10) for a precise definition. Then,

$$\hat{\boldsymbol{\beta}}_N^{\mathrm{LS}}(\lambda_j) = \underset{\boldsymbol{\beta}\in\mathbb{R}^2}{\mathrm{Arg\ min}} \sum_{i=1}^{N} (Y_i - c'_{Ni}(\lambda_j)\boldsymbol{\beta})^2 \ , \tag{3}$$

where

$$c'_{Ni}(\lambda_j) = (\cos(i\lambda_j) \ \sin(i\lambda_j)) \ . \tag{4}$$

The solution of (3) is

$$\hat{\boldsymbol{\beta}}_N^{\mathrm{LS}}(\lambda_j) = (C'C)^{-1}C'\mathbf{Y} \ , \tag{5}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, C and $C'C$ are defined by

$$C = \begin{pmatrix} \cos(\lambda_j) & \sin(\lambda_j) \\ \cos(2\lambda_j) & \sin(2\lambda_j) \\ \vdots & \vdots \\ \cos(N\lambda_j) & \sin(N\lambda_j) \end{pmatrix} \tag{6}$$

and

$$C'C = \begin{pmatrix} \sum_{k=1}^{N}\cos(k\lambda_j)^2 & \sum_{k=1}^{N}\cos(k\lambda_j)\sin(k\lambda_j) \\ \sum_{k=1}^{N}\cos(k\lambda_j)\sin(k\lambda_j) & \sum_{k=1}^{N}\sin(k\lambda_j)^2 \end{pmatrix} = \frac{N}{2}\,\mathrm{Id}_2 \tag{7}$$

where Id_2 is the identity matrix 2 by 2. Hence,

$$\hat{\boldsymbol{\beta}}_N^{\mathrm{LS}}(\lambda_j) = \frac{2}{N} C'\mathbf{Y} = \frac{2}{N} \left(\sum_{k=1}^{N} Y_k \cos(k\lambda_j) \quad \sum_{k=1}^{N} Y_k \sin(k\lambda_j) \right)' = (\hat{\beta}_N^{\mathrm{LS},(1)}(\lambda_j), \hat{\beta}_N^{\mathrm{LS},(2)}(\lambda_j))'. \quad (8)$$

In view of (1),

$$I_N(\lambda_j) = \frac{N}{8\pi} \|\hat{\boldsymbol{\beta}}_N^{\mathrm{LS}}(\lambda_j)\|^2 = \frac{N}{8\pi} \left((\hat{\beta}_N^{\mathrm{LS},(1)}(\lambda_j))^2 + (\hat{\beta}_N^{\mathrm{LS},(2)}(\lambda_j))^2 \right) =: I_N^{\mathrm{LS}}(\lambda_j) \,, \quad (9)$$

where $\|\cdot\|$ denotes the classical Euclidean norm and $\hat{\boldsymbol{\beta}}_N^{\mathrm{LS}}(\lambda_j) = (\hat{\beta}_N^{\mathrm{LS},(1)}(\lambda_j), \hat{\beta}_N^{\mathrm{LS},(2)}(\lambda_j))'$ is the least square estimates of $\boldsymbol{\beta'} = (\beta^{(1)}, \beta^{(2)})$ see, for example, Fajardo et al. (2018) and Reisen et al. (2017) and references therein. Note that $I_N(\lambda_j)$ (9) can be derived for different choices of ε_i, $i = 1, \ldots, N$.

It is supposed here that

$$\varepsilon_i = G(\eta_i) \,. \quad (10)$$

In (10), G is a non null real-valued and skew symmetric measurable function (*i.e.* $G(-x) = -G(x)$, for all x) and $(\eta_i)_{i\geq 1}$ is a stationary Gaussian process with zero mean and unit variance. Additional assumptions of $(\eta_i)_{i\geq 1}$ will be given in the sequel of the paper.

Let $\psi(.)$ be a function satisfying the following assumptions.

(**A1**) $0 < \mathbb{E}[\psi^2(\varepsilon_1)] < \infty$.

(**A2**) The function ψ is absolutely continuous with its almost everywhere derivative ψ' satisfying $\mathbb{E}[|\psi'(\varepsilon_1)|] < \infty$ and such that the function $z \mapsto \mathbb{E}[|\psi'(\varepsilon_1 - z) - \psi'(\varepsilon_1)|]$ is continuous at zero.

(**A3**) ψ is nondecreasing, $\mathbb{E}[\psi'(\varepsilon_1)] > 0$ and $\mathbb{E}[\psi'(\varepsilon_1)^2] < \infty$.

(**A4**) ψ is skew symmetric, *i.e.* $\psi(-x) = -\psi(x)$, for all x.

It is now introduced the M-periodogram based on the M-estimator $\hat{\boldsymbol{\beta}}_N^{\mathrm{M}}$ of the parameter $\boldsymbol{\beta}$ defined in Equation (2). The M-estimator $\hat{\boldsymbol{\beta}}_N^{\mathrm{M}} = (\hat{\beta}_N^{(1)}, \hat{\beta}_N^{(2)})'$ is defined as the solution (t_1, t_2) of

$$\sum_{i=1}^{N} \cos(i\lambda_j)\, \psi(Y_i - \cos(i\lambda_j) t_1) = 0 \text{ and } \sum_{i=1}^{N} \sin(i\lambda_j)\, \psi(Y_i - \sin(i\lambda_j) t_2) = 0. \quad (11)$$

$\hat{\beta}_N^{(1)}$ and $\hat{\beta}_N^{(2)}$ can be also seen as the minimizers with respect to t_1 and t_2, respectively, of

$$\left| \sum_{i=1}^{N} \cos(i\lambda_j)\, \psi(Y_i - \cos(i\lambda_j) t_1) \right| \text{ and } \left| \sum_{i=1}^{N} \sin(i\lambda_j)\, \psi(Y_i - \sin(i\lambda_j) t_2) \right|, \quad (12)$$

where ψ satisfies the same assumptions as in Koul and Surgailis (2000). By analogy to (9), the robust periodogram $I_N^M(\lambda_j)$ at $\lambda_j = 2\pi j/N, j = 1, \ldots, [N/2]$, is defined by

$$I_N^M(\lambda_j) = \frac{N}{8\pi} \|\hat{\boldsymbol{\beta}}_N^{\mathrm{M}}(\lambda_j)\|^2 = \frac{N}{8\pi} \left((\hat{\beta}_N^{(1)}(\lambda_j))^2 + (\hat{\beta}_N^{(2)}(\lambda_j))^2 \right) \,. \quad (13)$$

2.1.1 $M-$Periodogram in short-memory processes In this subsection the asymptotic properties of $\hat{\boldsymbol{\beta}}_N^{\mathrm{M}}$ are established in the short-range dependence framework. For this, the following assumptions are introduced. This result helps to establish the theoretical properties of the robust periodogram I_N^M given in Corollary 1.

(**A5**) Let η_t, $t \in \mathbb{Z}$, be i.i.d. standard Gaussian random variables and let a_j be real numbers such that $\sum_{j\geq 0} |a_j| < \infty$ and $a_0 = 1$. Then,

$$\varepsilon_i = \sum_{j\geq 0} a_j \eta_{i-j}.$$

(**A6**) ψ is the Huber function that is $\psi(x) = \max[\min(x, c), -c]$, for all x in $\mathbb{R}$, where c is a positive constant.

Theorem 1. *Assume that (A5) and (A6) hold and that $\boldsymbol{\beta} = \mathbf{0}$ in (2) so that $Y_i = \varepsilon_i$. Then, for any fixed j, $\hat{\boldsymbol{\beta}}_N^M$ defined by (12) satisfies*

$$\sqrt{\frac{N}{2}}(F(c) - F(-c))\hat{\boldsymbol{\beta}}_N^M(\lambda_j) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Delta}^{(j)}\right),\ N \to \infty ,$$

where F is the c.d.f. of ε_1 and

$$\boldsymbol{\Delta}^{(j)} = \sum_{k \in \mathbb{Z}} \mathbb{E}\{\psi(\varepsilon_0)\psi(\varepsilon_k)\} \begin{pmatrix} \cos(k\lambda_j) & \sin(k\lambda_j) \\ -\sin(k\lambda_j) & \cos(k\lambda_j) \end{pmatrix}.$$

Theorem 1 is proved in Section 5.

Corollary 1. *Under the assumptions of Theorem 1, $I_N^M(\lambda_j)$ defined in (13) satisfies for any fixed j,*

$$I_N^M(\lambda_j) \xrightarrow{d} \frac{X^2 + Y^2}{4\pi(F(c) - F(-c))^2}, \quad \text{as } N \to \infty ,$$

where

$$X \sim \mathcal{N}\left(0, \sum_{k \in \mathbb{Z}} \mathbb{E}\{\psi(\varepsilon_0)\psi(\varepsilon_k)\} \cos(k\lambda_j)\right),\ Y \sim \mathcal{N}\left(0, \sum_{k \in \mathbb{Z}} \mathbb{E}\{\psi(\varepsilon_0)\psi(\varepsilon_k)\} \cos(k\lambda_j)\right)$$

and

$$\mathrm{Cov}(X, Y) = \sum_{k \in \mathbb{Z}} \mathbb{E}\{\psi(\varepsilon_0)\psi(\varepsilon_k)\} \sin(k\lambda_j).$$

The proof of Corollary 1 is a straightforward consequence of Theorem 1 and (13).

2.1.2 *M*-periodogram for long-memory processes Now, consider the following assumption for $(\eta_i)_{i \geq 1}$ in the case of long-memory process. The results in this subsection are well detailed in Fajardo et al. (2018).

(**A7**) $(\eta_i)_{i \geq 1}$ is a stationary zero-mean Gaussian process with covariances $\rho(k) = \mathbb{E}(\eta_1 \eta_{k+1})$ satisfying:

$$\rho(0) = 1 \text{ and } \rho(k) = k^{-D} L(k),\ 0 < D < 1 ,$$

where the function L is slowly varying at infinity and is positive for large k. Recall that a slowly varying function $L(x)$, $x > 0$ is such that $L(xt)/L(x) \to 1$, as $x \to \infty$ for any $t > 0$. Constants and logarithms are example of slowly varying functions.
Moreover, the spectral density f of $(\eta_i)_{i \geq 1}$ can be expressed as:

$$f(\lambda) = |1 - \exp(-\mathrm{i}\lambda)|^{-2d} f^*(\lambda) , \tag{14}$$

where $d \in (0, 1/2)$ and f^* is an even, positive, continuous function on $(-\pi, \pi]$, bounded above and bounded away from zero.

Note that

$$D = 1 - 2d , \tag{15}$$

where D is defined in Assumption (A7) and d is the standard long-memory parameter notation given in the literature of long-memory models. The fact that $(\eta_i)_{i \geq 1}$ is required to satisfy (A7) essentially means that both $L(x)$, $x \geq 1$ and $f^*(\lambda)$, λ in $(-\pi, \pi]$ satisfy some smoothness properties.

Theorem 2. *Assume that (A7), (A1), (A2), (A3) and (A4) hold and that* $\boldsymbol{\beta} = 0$ *in (2) so that* $Y_i = \varepsilon_i$*. Then, for any fixed* j*,* $\hat{\boldsymbol{\beta}}_N^M(\lambda_j)$ *defined by (12) satisfies*

$$\sqrt{\frac{N}{2}}\hat{\boldsymbol{\beta}}_N^M(\lambda_j) = \frac{J_1}{\mathbb{E}[\psi'(\varepsilon_1)]}\left\{\sqrt{\frac{2}{N}}\sum_{i=1}^{N}\begin{pmatrix}\cos(i\lambda_j)\\ \sin(i\lambda_j)\end{pmatrix}\eta_i\right\} + o_p(N^{(1-D)/2})\ ,\ as\ N\to\infty\ , \tag{16}$$

where $J_1 = \mathbb{E}[\psi(G(\eta))\eta] \neq 0$*,* η *being a standard Gaussian random variable and* $D = 1-2d$*. Moreover,*

$$N^{D/2}\hat{\boldsymbol{\beta}}_N^M(\lambda_j) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{J_1^2}{(\mathbb{E}[\psi'(\varepsilon_1)])^2}\widetilde{\Gamma}\right)\ ,\ N\to\infty\ , \tag{17}$$

where

$$\widetilde{\Gamma} = \lim_{N\to\infty}\frac{4}{N^{2-D}}\sum_{1\le k,\ell\le N} c_{Nk}(\lambda_j)c_{N\ell}^T(\lambda_j)\rho(k-\ell) \tag{18}$$

$$= 8\pi\times(2\pi j)^{-2d}f^*(0)\begin{pmatrix}\mathcal{L}_1 & 0\\ 0 & \mathcal{L}_2\end{pmatrix}\ . \tag{19}$$

In Relation (18), the vector $c_{Nk}(\lambda_j)$ *is defined in (4),*

$$\mathcal{L}_1 = \frac{1}{\pi}\int_{\mathbb{R}}\frac{\sin^2(\lambda/2)}{(2\pi j-\lambda)^2}\left|\frac{\lambda}{2\pi j}\right|^{-2d}\mathrm{d}\lambda - \frac{1}{\pi}\int_{\mathbb{R}}\frac{\sin^2(\lambda/2)}{(2\pi j-\lambda)(2\pi j+\lambda)}\left|\frac{\lambda}{2\pi j}\right|^{-2d}\mathrm{d}\lambda\ , \tag{20}$$

and

$$\mathcal{L}_2 = \frac{1}{\pi}\int_{\mathbb{R}}\frac{\sin^2(\lambda/2)}{(2\pi j-\lambda)^2}\left|\frac{\lambda}{2\pi j}\right|^{-2d}\mathrm{d}\lambda + \frac{1}{\pi}\int_{\mathbb{R}}\frac{\sin^2(\lambda/2)}{(2\pi j-\lambda)(2\pi j+\lambda)}\left|\frac{\lambda}{2\pi j}\right|^{-2d}\mathrm{d}\lambda\ . \tag{21}$$

Corollary 2. *Under the assumptions of Theorem 2, the periodogram* I_N^M *defined in (13) satisfies*

$$N^{D-1}I_N^M(\lambda_j) \xrightarrow{d} (Z_1^2+Z_2^2)\ ,\ as\ N\to\infty\ , \tag{22}$$

where (Z_1, Z_2) *is a zero-mean uncorrelated Gaussian vector with covariance matrix equal to*

$$\frac{J_1^2}{8\pi(\mathbb{E}[\psi'(\varepsilon_1)])^2}\widetilde{\Gamma}\ , \tag{23}$$

with $\widetilde{\Gamma}$ *defined in (18).*

Theorem 2 and Corollary 2 are proved in Fajardo et al. (2018).

2.2 Q_N-periodogram

Another possible approach to obtain the classical periodogram (1) is to write it in terms of the sample autocovariance function

$$I_N(\lambda_j) = \frac{1}{2\pi}\sum_{h=-(N-1)}^{N-1}\widehat{\gamma}(h)\cos(h\lambda_j), \tag{24}$$

where $\lambda_j = 2\pi j/N, j = 1,\dots,[N/2]$ and $\widehat{\gamma}(h)$ is the classical sample autocovariance function for a sample $\{Y_1,...,Y_N\}$.

A straightforward approach to robustify 24 is to plug in a robust autocovariance function replacing the classical one. This methodology is now addressed.

For a sample $x_1,...,x_N$ Rousseeuw and Croux (1993) proposed a robust scale estimator function $Q_N(\cdot)$ which is based on the τth order statistic of $\binom{N}{2}$ distances $\{|x_j-x_k|, j<k\}$, and can be written as

$$Q_N(x) = \kappa\times\{|x_j-x_k|; j<k\}_{(\tau)}, \tag{25}$$

where κ is a constant used to guarantee consistency ($\kappa = 2.2191$ for the Gaussian distribution) and $\tau = \lfloor(\binom{N}{2}+2)/4\rfloor + 1$. The above function can be evaluated using the algorithm proposed by Croux and Rousseeuw (1992), which is computationally efficient.

Based on $Q_N(\cdot)$, Ma and Genton (2000) proposed a highly robust estimator for the autocovariance function:

$$\widehat{\gamma}_{Q_N}(h) = \frac{1}{4}\left[Q^2_{N-h}(\mathbf{u}+\mathbf{v}) - Q^2_{N-h}(\mathbf{u}-\mathbf{v})\right], \tag{26}$$

where $\mathbf{u}$ and $\mathbf{v}$ are vectors containing the initial $N-h$ and the final $N-h$ observations of $x_1, ..., x_N$, respectively. The robust estimator for the autocorrelation function is

$$\widehat{\rho}_{Q_N}(h) = \frac{Q^2_{N-h}(\mathbf{u}+\mathbf{v}) - Q^2_{N-h}(\mathbf{u}-\mathbf{v})}{Q^2_{N-h}(\mathbf{u}+\mathbf{v}) + Q^2_{N-h}(\mathbf{u}-\mathbf{v})}. \tag{27}$$

It can be shown that $|\widehat{\rho}_{Q_N}(h)| \leq 1$ for all h.

Now, returning to (24), the robust Q_N-periodogram for a sample $\{Y_1, ..., Y_N\}$ is defined by

$$I_N^{Q_N}(\lambda_j) = \frac{1}{2\pi}\sum_{h=-(N-1)}^{N-1} \widehat{\gamma}_{Q_N}(h)\cos(h\lambda_j), \tag{28}$$

where $\lambda_j = 2\pi j/N, j = 1, \ldots, [N/2]$.

The theoretical properties of $I_N^{Q_N}$ are still under study. Therefore, in the sequel, the asymptotic properties of $\hat{\gamma}_{Q_N}$ are summarized for short and long memory processes. These are well detailed in Lévy-Leduc et al. (2011).

2.2.1 Main asymptotic results for short memory process In the short-memory scenario, the process under study $(Y_i)_{i\geq 1}$ satisfies the following assumption (see, also, Lévy-Leduc et al. (2011)):

(A8) $(Y_i)_{i\geq 1}$ is a stationary zero-mean Gaussian process with autocovariance sequence $\gamma(h) = \mathbb{E}(Y_1 Y_{h+1})$ satisfying:

$$\sum_{h\geq 1}|\gamma(h)| < \infty\ .$$

Theorem 3. *Assume that (A8) holds and let h be a non negative integer. Then, the autocovariance estimator $\widehat{\gamma}_{Q_N}(h)$ satisfies the following Central Limit Theorem:*

$$\sqrt{N}\,(\widehat{\gamma}_{Q_N}(h) - \gamma(h)) \xrightarrow{d} \mathcal{N}(0, \check{\sigma}^2_h), N \to \infty,$$

where

$$\check{\sigma}^2(h) = \mathbb{E}[\zeta^2(Y_1, Y_{1+h})] + 2\sum_{k\geq 1}\mathbb{E}[\zeta(Y_1, X_{1+h})\zeta(Y_{k+1}, Y_{k+1+h})]\ , \tag{29}$$

and the function ζ is defined by

$$\zeta : (x,y) \mapsto \left\{(\gamma(0)+\gamma(h))\ \mathrm{IF}\left(\frac{x+y}{\sqrt{2(\gamma(0)+\gamma(h))}}, Q, \Phi\right) - (\gamma(0)-\gamma(h))\ \mathrm{IF}\left(\frac{x-y}{\sqrt{2(\gamma(0)-\gamma(h))}}, Q, \Phi\right)\right\}. \tag{30}$$

where IF *is defined by*

$$\mathrm{IF}(x, Q, \Phi) = \kappa\left(\frac{1/4 - \Phi(x+1/\kappa) + \Phi(x-1/\kappa)}{\int_{\mathbb{R}}\phi(y)\phi(y+1/\kappa)\mathrm{d}y}\right), \tag{31}$$

where Φ and ϕ denote the c.d.f. and p.d.f. of a standard Gaussian random variable, respectively with κ defined in (25).

Theorem 3 is proved in Lévy-Leduc et al. (2011).

2.2.1 Main asymptotic results for long-memory process The following results concern the robust autocovariance function for long-memory process see, also, Lévy-Leduc et al. (2011).

(**A9**) $(Y_i)_{i\geq 1}$ is a stationary zero-mean Gaussian process with autocovariance $\gamma(h) = \mathbb{E}(Y_1 Y_{h+1})$ satisfying:

$$\gamma(h) = h^{-D} L(h),\ 0 < D < 1\ ,$$

where L is slowly varying at infinity and is positive for large h. Note that, as previously stated, $D = 1 - 2d$.

Theorem 4. *Assume that (A9) holds and that L has three continuous derivatives. Assume also that $L_i(x) = x^i L^{(i)}(x)$ satisfy: $L_i(x)/x^{\epsilon} = O(1)$, for some ϵ in $(0, D)$, as x tends to infinity, for all $i = 0, 1, 2, 3$, where $L^{(i)}$ denotes the ith derivative of L. Let h be a non negative integer. Then, $\widehat{\gamma}_{Q_N}(h)$ satisfies the following limit theorems as N tends to infinity.*

(i) If $D > 1/2$,

$$\sqrt{N}\,(\widehat{\gamma}_{Q_N}(h) - \gamma(h)) \xrightarrow{d} \mathcal{N}(0, \check{\sigma}^2(h))\ ,$$

where

$$\check{\sigma}^2(h) = \mathbb{E}[\zeta^2(Y_1, Y_{1+h})] + 2\sum_{k\geq 1} \mathbb{E}[\zeta(Y_1, Y_{1+h})\zeta(Y_{k+1}, Y_{k+1+h})]\ ,$$

ζ being defined in (30).

(ii) If $D < 1/2$,

$$\beta(D)\frac{N^D}{\widetilde{L}(N)}\,(\widehat{\gamma}_{Q_N}(h) - \gamma(h)) \xrightarrow{d} \frac{\gamma(0) + \gamma(h)}{2}(Z_{2,D}(1) - Z_{1,D}(1)^2)$$

where $\beta(D) = \mathrm{B}((1 - D)/2, D)$, B denotes the Beta function, the processes $Z_{1,D}(\cdot)$ and $Z_{2,D}(\cdot)$ are defined as follows:

$$Z_{1,D}(t) = \int_{\mathbb{R}} \left[\int_0^t (u - x)_+^{-(D+1)/2} \mathrm{d}u\right] \mathrm{d}B(x),\quad 0 < D < 1\ , \tag{32}$$

$$Z_{2,D}(t) = \int'_{\mathbb{R}^2} \left[\int_0^t (u - x)_+^{-(D+1)/2} (u - y)_+^{-(D+1)/2} du\right] \mathrm{d}B(x)\mathrm{d}B(y),\ 0 < D < 1/2\ , \tag{33}$$

and

$$\widetilde{L}(N) = 2L(N) + L(N + h)(1 + h/N)^{-D} + L(N - h)(1 - h/N)^{-D}\ , \tag{34}$$

where B is the standard Brownian motion. The symbol $\int'$ means that the domain of integration excludes the diagonal.

Theorem 4 is proved in Lévy-Leduc et al. (2011).

3 Monte Carlo simulation

In this section, small sample size experiments are conducted with the aim to clarify the empirical performance of the spectral estimates discussed previously in a different context such as time series with additive outliers. Based on this, some standard questions, such as (1) what is the best method to be used in a real application? (2) which method (if any) should be considered when dealing with outliers? (3) Does the large observation (if any) make similar outlier´s effect on the statistical time series modelling functions, that is, on the ACF and periodogram functions? among others, are expected to be answered or, at least, clarified.

Let $\{X_t\}_{t=1,\ldots,N}$ be a sample from a Gaussian second order stationary process and let $\{Y_t\}_{t=1,\ldots,N}$ be a sample of the process defined by

$$Y_t = X_t + \omega W_t \tag{35}$$

where the parameter ω represents the magnitude of the outlier, and W_t is a random variable with probability distribution

$$\mathbb{P}(W_t = -1) = \mathbb{P}(W_t = 1) = \delta/2 \text{ and } \mathbb{P}(W_t = 0) = 1 - \delta ,$$

where $\mathbb{E}[W_t] = 0$ and $\mathbb{E}[W_t^2] = \mathrm{Var}(W_t) = \delta$. Note that (35) is based on the parametric models proposed by Fox (1972). W_t is the product of *Bernoulli*(δ) and *Rademacher* random variables; the latter equals 1 or -1, both with probability $1/2$. X_t and W_t are independent random variables. Note that, if $\omega = 0.0$ $\{Y_t\}$ is an outlier free time series.

In order to compare the performance of M- and Q_N-periodogram, a Monte Carlo investigation was carried out under different contamination scenarios. For the simulations, the number of replications was 5000, the samples $\{X_t\}$ of size $N = 500$ were generated according to a model autocorrelation structure, which is given in what follows, and the contaminated data Y_t were generated from (35) with $\delta = 0.01$ for magnitudes $\omega = 0$ (no outliers) and 10.

The comparison between the methods is performed by estimating α in the linear regression $\log(I(\lambda_j)) \simeq const + \alpha \log(\lambda_j) + E_j$, $j = 1, \ldots, N^{0.7}$, where $I(.)$ is either $I_N(.)$, $I_N^M(.)$ or $I_N^{Q_N}(.)$. The data were generated based on

$$X_t = (1-B)^{-d} Z_t = \sum_{j \geq 0} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \epsilon_{t-j} , \tag{36}$$

where ϵ_t is an AR(1) model, that is, $\epsilon_t = \phi \epsilon_{t-1} + \eta_t$, where η_t, $t = 1, ..., N$, are i.i.d. standard Gaussian random variables.

In the finite sample size investigation, the model correlation structures are divided in two cases:

1. An AR(1) model with $\phi = 0.6$ and $d = 0$.
2. An ARFIMA$(0, d, 0)$ model with $d = 0.3$.

Figure 1 displays the plots of the empirical densities of $\hat{\alpha}_{I_N}$, $\hat{\alpha}_{I_N^M}$ and $\hat{\alpha}_{I_N^{Q_N}}$ for the case of AR(1) models without contamination ($\omega = 0$). Although, $\hat{\alpha}_{I_N^M}$ has a slight better performance than $\hat{\alpha}_{I_N^{Q_N}}$, that is, the first method and the classical periodogram presented very close densities, all the methods provided similar results showing that, even for small sample sizes, the empirical density is very close which corroborate the theoretical results discussed previously. Based on the asymptotic theory and the empirical results all three methods can be used to estimate the spectral density of a time series when there is no contamination of additive outliers. This opens an important contribution in the context that alternative spectral estimators such as I_N^M and $I_N^{Q_N}$ can be used instead of the classical periodogram I_N in the step procedure for modelling time series data. For example, these estimators can be an alternative tools to be used in the Whittle function to obtain the parameter estimates. This will be also discussed in what follows. Note that, the disadvantage of $I_N^{Q_N}$ over I_N^M and I_N is that the ACF using $Q_N(.)$ does not have the positive definite property.

When the data is contaminated with additive outliers the scenario changes significantly. As well known, the periodogram, which depends on the classical autocovariance, is corrupted by the outliers. Therefore, the alternative methods are almost unaffected. This is displayed in Figure 2 in which $\omega = 10$ and $\delta = 0.01$. The empirical density of $\hat{\alpha}_{I_N}$ is shifted to the right side which is an expected result since the variance increases with outliers. The empirical densities of $\hat{\alpha}_{I_N^M}$ and $\hat{\alpha}_{I_N^{Q_N}}$ remain almost unchangeable.

In the case of long-memory process, the empirical density plots are given in Figures 3 and 4 for non-contaminated and contaminated time series, respectively. Similar conclusions of the AR case

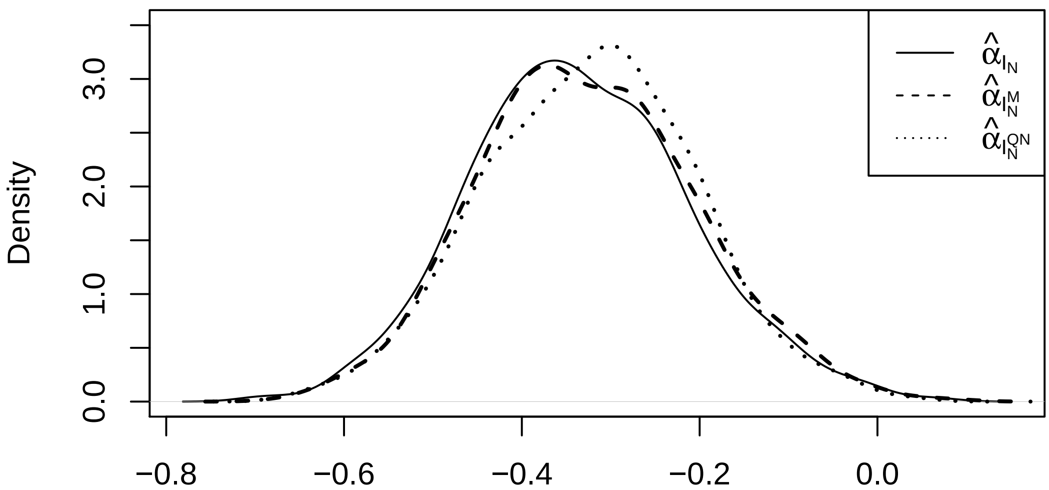

Fig. 1. Densities of $\hat{\alpha}_{I_N}$, $\hat{\alpha}_{I_N^M}$ and $\hat{\alpha}_{I_N^{QN}}$ for AR(1) models with $\phi = 0.6$ and $\omega = 0$.

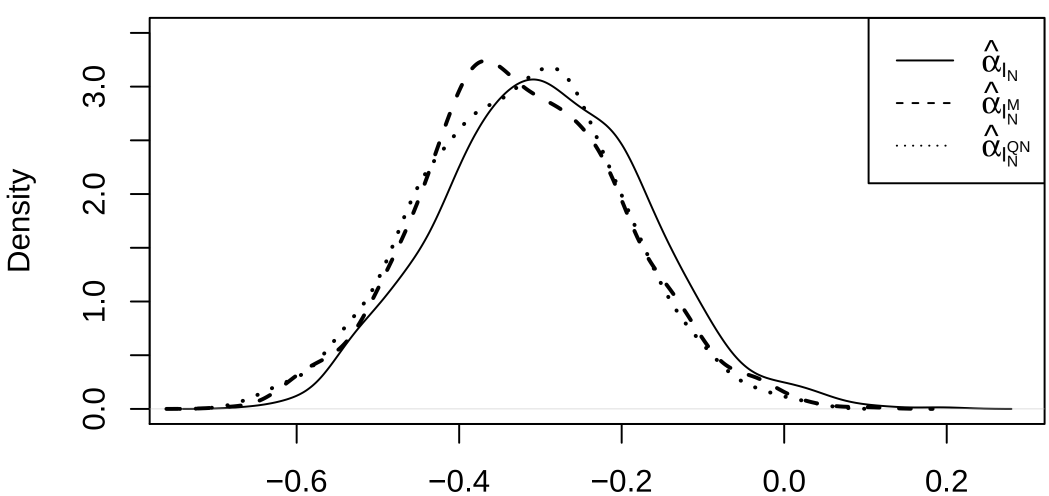

Fig. 2. Densities of $\hat{\alpha}_{I_N}$, $\hat{\alpha}_{I_N^M}$ and $\hat{\alpha}_{I_N^{QN}}$ for AR(1) models with $\phi = 0.6$, $\delta = 0.01$ and $\omega = 10$.

are drawn. That is, in the uncontaminated scenarios, all three methods displayed similar densities although the method M and the classical one (periodogram) are very close. In the contaminated case, the classical one is totally affected by the additive outliers. Reinforcing that the ACF using Q_N does not have the positiveness property.

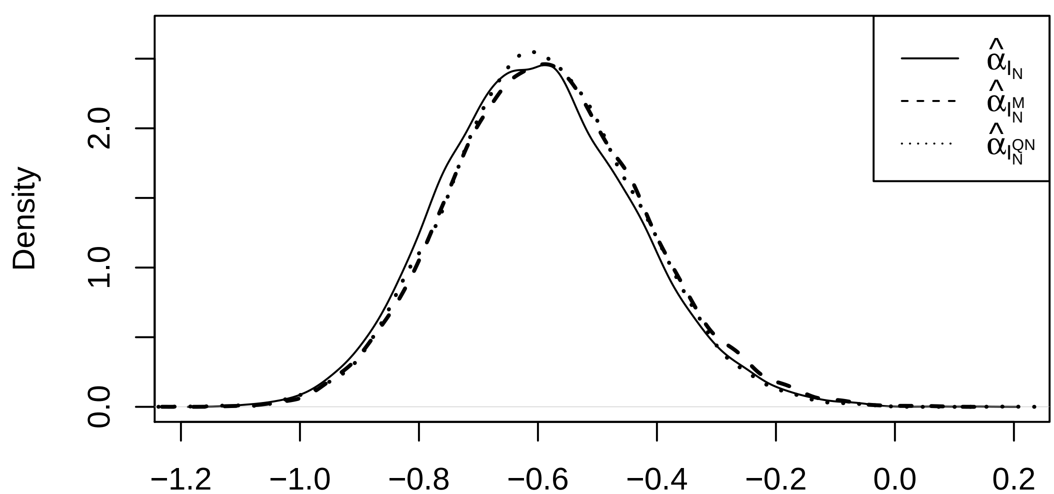

Fig. 3. Densities of $\hat{\alpha}_{I_N}$, $\hat{\alpha}_{I_N^M}$ and $\hat{\alpha}_{I_N^{Q_N}}$ when $d = 0.3$, $N = 500$ and $\omega = 0$.

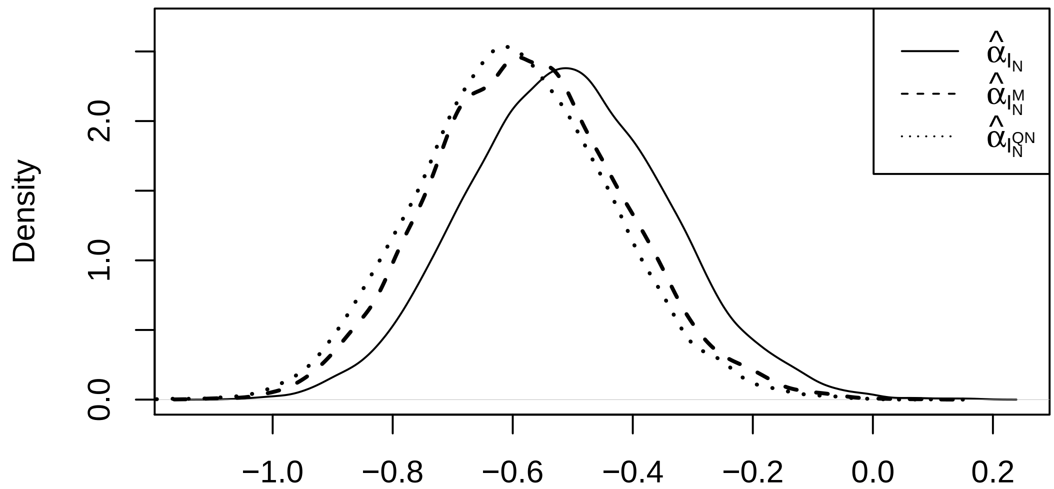

Fig. 4. Densities of $\hat{\alpha}_{I_N}$, $\hat{\alpha}_{I_N^M}$ and $\hat{\alpha}_{I_N^{Q_N}}$ when $d = 0.3$, $N = 500$, $\delta = 0.01$ and $\omega = 10$.

4 Applications of M and Q_N-periodograms

4.1 Robust estimation of the fractional parameter

Based on the theoretical results discussed previously, this section introduces some applications related to the use of $M-$regression and Q_N estimation functions. The application is divided in two cases: (a) Estimation of the fractional parameter d in long-memory processes; (b) Estimation in periodic AR (PAR) processes. Some finite sample size investigation is also addressed in the context of time series with and without outliers.

(a) *Estimation of the fractional parameter in long-memory process*

The estimation methods of the fractional parameter d discussed here are derived from the well-known semi-parametric regression method (GPH) originally proposed by Geweke and Porter-Hudak (1983). The regression estimation methods based on I_N^M and $I_N^{Q_N}$ were previously introduced in Reisen et al. (2017) and Fajardo et al. (2009), respectively, papers where the reader will find more details related to theoretical and empirical results of these estimation methodologies.

(**A10**) $(\varepsilon_i)_{i\geq 1}$ is a stationary mean-zero Gaussian process with spectral density given in Assumption (A7).

For estimating the fractional parameter d of long-memory processes having their spectral density satisfying (14), it is usual to use the standard GPH (Geweke and Porter-Hudak (1983)) estimator defined in the following. This estimator is motivated heuristically by starting from

$$\log(f(\lambda_j)) = -2d\log(|2\sin(\lambda_j/2)|) + \log(f^*(\lambda_j)) = -2dX_j + \log(f^*(\lambda_j)) \\ = \log(f_0^\star) - 2dX_j + \log(f_j^\star/f_0^\star), \quad (37)$$

where $X_j = \log|2\sin(\lambda_j/2)|$ and $f_j^\star = f^\star(\lambda_j)$. If

$$\varepsilon_j^R = \log\left(\frac{I_N(\lambda_j)}{f(\lambda_j)}\right), \tag{38}$$

then

$$\log(I_N(\lambda_j)) = \varepsilon_j^R + \log(f(\lambda_j)),$$

and, by (37),

$$\log(I_N(\lambda_j)) = \log(f_0^\star) - 2dX_j + \log(f_j^\star/f_0^\star) + \varepsilon_j^R. \tag{39}$$

The GPH estimator is given by

$$\hat{d}^{\mathrm{GPH}} = \frac{-0.5\sum_{j=1}^{m_N}(X_j - \bar{X})\log(I_N^{\mathrm{LS}}(\lambda_j))}{\sum_{k=1}^{m_N}(X_k - \bar{X})^2}, \tag{40}$$

where $X_j = \log|2\sin(\lambda_j/2)|$, $\bar{X} = \sum_{j=1}^{m_N} X_j/m_N$, $I_N^{\mathrm{LS}}(\lambda_j)$ is defined in (9) and m_N is a function of N.

Based on the above discussion, one way to define a $M-$regression estimator of d consists in replacing I_N^{LS} in (40) by I_N^M defined in (13):

$$\hat{d}^M = \frac{-0.5\sum_{j=1}^{m_N}(X_j - \bar{X})\log(I_N^M(\lambda_j))}{\sum_{k=1}^{m_N}(X_k - \bar{X})^2}, \tag{41}$$

where $X_j = \log|2\sin(\lambda_j/2)|$, $\bar{X} = \sum_{j=1}^{m_N} X_j/m_N$ and m_N is a function of N which is specified in Theorem 5.

The theoretical properties of $\hat{d}^{\mathrm{M}}$ are established under the following assumptions. The random process (ε_j) is obtained through a moving average process:

$$\varepsilon_j = \sum_{k\le j} a_{j-k}\zeta_k\ , \quad a_j = L(j)j^{-(1+D)/2}\ ,\ j \ge 1\ , \tag{42}$$

for some D in $(0,1)$, where $L(\cdot)$ is a positive slowly varying function at infinity and where the random variables ζ_k are i.i.d. with zero mean and variance 1. It is assumed that the distribution of ζ_0 satisfies

$$\left|\mathbb{E}(\mathrm{e}^{\mathrm{i}u\zeta_0})\right| \le C(1+|u|)^{-\delta}\ ,\ u \in \mathbb{R}\ . \tag{43}$$

where $C < \infty$ and $\delta > 0$ are constants. Note that, Conditions (42) and (43) imply that the cumulative distribution function F_{ε_0} of ε_0 is infinitely boundedly differentiable, see Koul and Surgailis (2000).

Theorem 5. *Let $Y_i = \varepsilon_i$, for all i in $\{1,\ldots,N\}$, where ε_i satisfy (42) and (A10). Assume that $1/D$ is not an integer and that $\boldsymbol{\beta} = \mathbf{0}$ in (2). Assume moreover that $\mathbb{E}(\zeta_0^{4\vee 2k^\star}) < \infty$, where $k^\star = [1/D]$, ζ_0 is defined in (42) and satisfies (43), $\nu_1 \neq 0$, $\nu_2 = 0$ and $\nu_3 \neq 0$, where the ν_k are defined by*

$$\nu_k = \int_0^\infty \psi(y)\left[1-(-1)^k\right] f^{(k)}(y)\mathrm{d}y, \ \ \textit{for all integer } k \ge 0\ , \tag{44}$$

where ψ is the Huber function. Then, if $1/3 < D < 1$,

$$\sqrt{m_N}(\hat{d}^M - d) \xrightarrow{d} \mathcal{N}(0, \pi^2/24), \ \textit{as } N \to \infty, \tag{45}$$

where $\hat{d}^M$ is defined in (41) and $m_N = N^\beta$ with $0 < \beta < (1-D)/3$.

This result is proved in Reisen et al. (2017).

Another way of defining a robust estimator of d is to consider:

$$\hat{d}^{Q_N} = \frac{-0.5\sum_{j=1}^{m_N}(X_j - \bar{X})\log(I_N^{Q_N}(\lambda_j))}{\sum_{k=1}^{m_N}(X_k - \bar{X})^2}, \tag{46}$$

where $X_j = \log|2\sin(\lambda_j/2)|$, $\bar{X} = \sum_{j=1}^{m_N} X_j/m_N$, $I_N^{Q_N}(\lambda_j)$ is defined in (28) and m_N is a function of N. For further information, see Fajardo et al. (2009). The asymptotic property of $\hat{d}^{Q_N}$ is still an open problem, however, the empirical results given in Fajardo et al. (2009) support the use of this method under time series with and without outliers. The performance of fractional estimators $\hat{d}^{GPH}$, $\hat{d}^{M}$ and $\hat{d}^{Q_N}$ is the motivation of the next subsection for long-memory time series with and without additive outliers.

4.1.1- Finite sample size investigation

In this subsection, the numerical experiments were carried out in accordance with the model of Section 3. For the simulations, $N = 500$, $\omega = 10$ and $\delta = 0.01$ for 5000 replications. The results are displayed in Figures 5, 6 and Table 1. Since there is not short-memory component in the model m_N was fixed at $N^{0.7}$ for all tree methods.

Figure 5 presents the boxplots with the results of $\hat{d}_{GPH}$, $\hat{d}_M$ and $\hat{d}_{Q_N}$ estimators for the uncontaminated scenario. $\hat{d}_M$ and $\hat{d}_{Q_N}$ seem to present positive bias and, surprisingly, $\hat{d}_{Q_N}$ displays smaller deviation. However, in general, all methods perform similarly, i.e., all estimation methods leaded to comparable estimates close to the real values of d.

Figure 6 displays the boxplots of $\hat{d}_{GPH}$, $\hat{d}_M$ and $\hat{d}_{Q_N}$ when the series has outliers. As can be perceived from the boxplots, the GPH estimator is clearly affected by additive outliers while the robust ones keep almost the same picture as the one of the non-contaminated scenario, except that the bias of $\hat{d}_{Q_N}$ becomes negative, that is, this estimator tends to overestimate the true parameter.

The empirical mean, bias and mean square root are displayed in Table 1. This numerically corroborates the results discussed based on Figures 5, 6, that is, the estimators have similar performance in the absence of outliers in the data. While the performance of $\hat{d}_{GPH}$ changes dramatically in the presence of outliers, the estimates from $\hat{d}_{Q_N}$ and $\hat{d}_M$ keep almost unchangeable. As a general conclusion, the empirical result suggests that all the methods can be used to estimate the parameter d when there is not a suspicion of additive or abrupt observation. However, in the existence of a single atypical observation, the methods $\hat{d}_{Q_N}$ and $\hat{d}_M$ should be preferred. Similar conclusions are given in Fajardo et al. (2009) and Reisen et al. (2017) for $\hat{d}_{Q_N}$ and $\hat{d}_M$, respectively.

Table 1. Empirical Mean, Bias and RMSE of $\hat{d}_{GPH}$, $\hat{d}_M$ and $\hat{d}_{Q_N}$ when $\omega = 10$ and $\delta = 0, 0.01, 0.05$.

d	δ	MEAN			BIAS			RMSE		
		$\hat{d}_{GPH}$	$\hat{d}_M$	$\hat{d}_{Q_N}$	$\hat{d}_{GPH}$	$\hat{d}_M$	$\hat{d}_{Q_N}$	$\hat{d}_{GPH}$	$\hat{d}_M$	$\hat{d}_{Q_N}$
0.3	0.0	0.3029	0.2950	0.2933	0.0029	-0.0049	-0.0066	0.0601	0.0596	0.0558
	0.01	0.2226	0.2899	0.3052	-0.0773	-0.0101	0.0052	0.0972	0.0581	0.0584
	0.05	0.1225	0.2681	0.3236	-0.1775	-0.0318	0.0236	0.1873	0.0689	0.0682

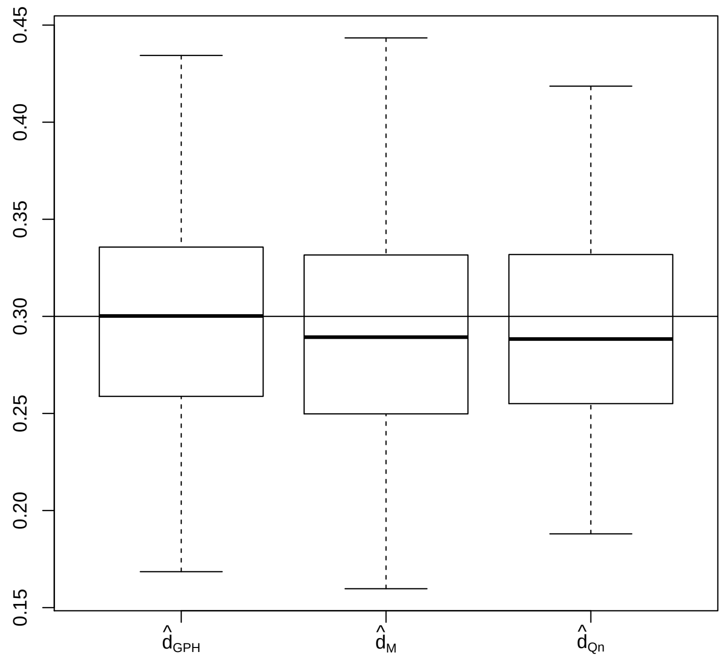

Fig. 5. Boxplots of $\hat{d}_{GPH}$, $\hat{d}_M$ and $\hat{d}_{Q_N}$ when $\delta = 0$.

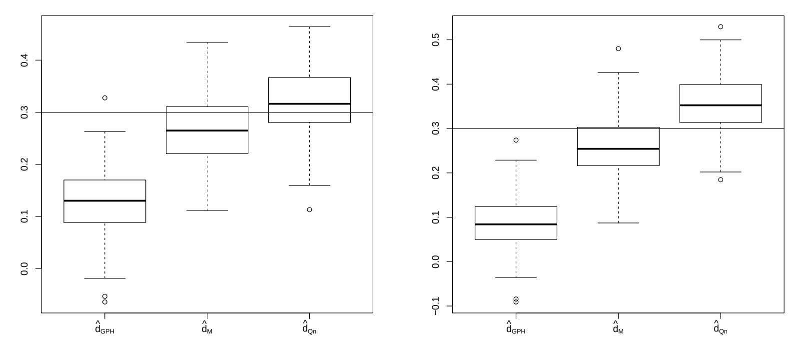

Fig. 6. Boxplots of $\hat{d}_{GPH}$, $\hat{d}_M$ and $\hat{d}_{Q_N}$ when $\delta = 0.05$ and $\delta = 0.1$, respectively.

4.2 Q_n and M-estimators in PARMA models

One of the most popular periodic causal process is the PARMA model which generalizes the ARMA model. $\{Z_t\}_{t\in\mathbb{Z}}$ is said to be a PARMA model if it satisfies the difference equation

$$\sum_{j=0}^{p_\nu} \phi_{\nu,j} Z_{r\mathcal{S}+\nu-j} = \sum_{k=0}^{q_\nu} \theta_{\nu,k} \varepsilon_{r\mathcal{S}+\nu-k}, r \in \mathbb{Z} \tag{47}$$

where for each season ν ($1 \leq \nu \leq \mathcal{S}$) where $\mathcal{S}$ is the period, p_ν and q_ν are the AR and MA orders, respectively, $\phi_{\nu,1}, \ldots, \phi_{\nu,p_\nu}$ and $\theta_{\nu,1}, \ldots, \theta_{\nu,q_\nu}$ are the AR and MA coefficients, respectively, and $\phi_{\nu,0} = \theta_{\nu,0} = 1$. The sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is zero-mean and uncorrelated, and has periodic variances with period $\mathcal{S}$, i.e. $\mathrm{E}(\varepsilon^2_{r\mathcal{S}+\nu}) = \sigma^2_\nu$ for $\nu = 1, \ldots, \mathcal{S}$. In the following, $p = \max_\nu p_\nu$, $q = \max_\nu q_\nu$, $\phi_{\nu,j} = 0$ for $j > p_\nu$, $\theta_{\nu,k} = 0$ for $k > q_\nu$, and (47) is referred as the PARMA$(p,q)_\mathcal{S}$ model (see, for example, Basawa and Lund (2001) and Sarnaglia et al. (2015)).

To deal with outliers effect in the estimation of PAR model, Sarnaglia et al. (2010) proposed the use of the $Q_N(.)$ function in this model. Following the same lines of the linear time series model described previously, the $Q_N(.)$ function is used to compute an estimator of the periodic autocovariance function $\gamma^{(\nu)}(h)$ at lag h and this sample ACF based on $Q_N(.)$ estimator, denoted here as $\gamma^{(\nu)}_Q(h)$, replaces the classical periodic ACF $\gamma^{(\nu)}(h)$ in the Yule-Walker periodic equations (see, for example, McLeod (1994) and Sarnaglia et al. (2010)) to derive an alternative parameter estimator method for a periodic AR model. The authors derived some asymptotic and empirical properties of the proposed estimator. They showed that the method well accommodate the effect of additive outliers, that is, it presented robustness against these type of observations in the finite sample size series as well as in a real data set.

Let now $Z_1, ...Z_N$, where $N = n\mathcal{S}$, be a sample from PAR process which is a particular case of the model definition in (47) with $q_\nu = 0$ and let now $Q_N(.)$ for PAR process be defined as

$$Q^{(\nu)}_N(Z) = Q_N(\{Z_{r\mathcal{S}+\nu}\}_{0\le r\le N}) \ . \tag{48}$$

Based on $Q^{(\nu)}_N(Z)$, the authors derived the sample ACF for periodic stationary processes $\hat{\gamma}^{(\nu)}_Q(h)$. Under some model assumptions, they proved the following main results.

1. For a fixed lag h, $\hat{\gamma}^{(\nu)}_Q(h)$ satisfies the following central limit theorem: As $N \longrightarrow \infty$,

$$\sqrt{N}\left(\hat{\gamma}^{(\nu)}_Q(h) - \gamma^{(\nu)}(h)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \check{\sigma}^2_h) \ ,$$

   where $\gamma^{(\nu)}(h)$ is the periodic ACF function and $\check{\sigma}^2_h$ is the variance, more details are given in Sarnaglia et al. (2010).
2. The $Q^{(\nu)}_N$ Yule-Walker estimators $(\tilde{\phi}_{\nu,i})_{1\le i\le p_\nu, \nu=1,\ldots,\mathcal{S}}$ satisfy $\tilde{\phi}_{\nu,i} - \phi_{\nu,i} = O_P(N^{-1/2})$ for all $i = 1, \ldots, p_\nu$ and ν in $\{1, \ldots, \mathcal{S}\}$.

Recently, Solci et al. (2018) compared the Yule-Walker estimator (YWE), the robust least squares estimator (Shao (2008)) and the ACF Q_n estimator ($\hat{\gamma}^{(\nu)}_Q(h)$, denoted here RYWE, in the context of estimating the parameters in PAR models with and without outliers. Their main conclusion is similar to the cases discussed previously, that is, for the case of ARFIMA model $\hat{\gamma}^{(\nu)}_Q(h)$ displayed good performance in estimating the parameters in PAR models, periodic samples with and without outliers. As expected, the YWE estimator performed very poorly with the presence of outliers in the data. One of their simulation results is reproduced in the table below (Table 2) in which $n = 100, 400$ (cycles), $\mathcal{S} = 4$, ϵ_t is a Gaussian white noise process and $\delta =$ 0.01 (outlier's probability) and magnitude $\omega = 10$. The results correspond to the mean of 5000 replications.

Table 2. Bias and RMSE for Model 1 and outliers with probability $\delta = 0.01$.

ω	ϵ_t	n	$\phi_{\nu,1}$	YWE Bias	YWE RMSE	RYWE Bias	RYWE RMSE
0	$\mathcal{N}(0,1)$	100	0.9	-0.007	0.077	-0.003	0.103
			0.8	-0.002	0.065	0.004	0.084
			0.7	0.000	0.063	-0.001	0.083
			0.6	-0.005	0.066	-0.003	0.083
		400	0.9	-0.001	0.037	-0.001	0.047
			0.8	-0.001	0.031	0.000	0.038
			0.7	-0.001	0.032	0.001	0.038
			0.6	0.000	0.032	0.000	0.039
7	$\mathcal{N}(0,1)$	100	0.9	-0.181	0.247	0.014	0.120
			0.8	-0.118	0.176	0.012	0.096
			0.7	-0.105	0.157	0.015	0.091
			0.6	-0.097	0.151	0.012	0.091
		400	0.9	-0.183	0.203	0.017	0.055
			0.8	-0.129	0.144	0.012	0.046
			0.7	-0.108	0.124	0.013	0.044
			0.6	-0.103	0.119	0.014	0.043

As an alternative estimator of $\tilde{\phi}_{\nu,i}$, Sarnaglia et al. (2016) proposed the use of $M-$periodogram function to obtain estimates of the parameters in PARMA models. The estimator is based on the approximated Whittle function suggested in Sarnaglia et al. (2015). Basically, the Whittle $M-$estimator of PARMA parameters is derived by the ordinary Fourier transform with the non-linear M-regression estimator for periodic processes in the harmonic regression equation that leads to the classical periodogram. The empirical simulation investigation in Sarnaglia et al. (2016) considered the scenarios of periodic time series with presence and absence of additive outliers. Their small sample size investigation leaded to a very promising estimation method under the context of modelling periodic time series with additive outliers and heavy-tailed distributions. The theoretical justification of the proposed estimator is still an open problem and it is now a current research theme of the authors.

Table 3 displays results of a simple simulation example to show the empirical performance of the Whittle $M-$estimator with the Huber function $\psi(x)$ (Huber (1964)) compared to the maximum Gaussian and Whittle likelihood estimators to estimate a PAR(2) model with parameters $\phi_{1,1} = -0.2$, $\phi_{2,1} = -0.5$, $\sigma^2_{1,1} = 1.0$ and $\sigma^2_{2,1} = 1.0$. The sample sizes are $N = n\mathcal{S} = 300$, 800 ($n = 150$, 400, respectively) and the Huber function was used with constant equal to 1.345, which ensure that the M-estimator is 95% as efficient as the least squares estimator for univariate multiple linear models with independent and identically distributed Gaussian white noise. The sample root mean square error (RMSE) was computed over 5000 replications. The PAR(2) model with additive outliers was generated with outlier's probability $\delta = 0.01$ and magnitude $\omega = 10$. The values with "$*$" refer to the RMSE for the contaminated series.

Table 3. Empirical RMSE results for estimating an PAR(2) model.

Method	N	$\phi_{1,1}$	$\sigma^2_{1,1}$	$\phi_{2,1}$	$\sigma^2_{2,1}$
MLE	300	0.067; 0.121*	0.117; 1.366*	0.079; 0.252*	0.111; 1.363*
	800	0.048; 0.101*	0.079; 1.122*	0.046; 0.239*	0.074; 1.253*
WLE	300	0.068; 0.121*	0.117; 1.368*	0.079; 0.252*	0.111; 1.364*
	800	0.048; 0.101*	0.079; 1.122*	0.046; 0.239*	0.074; 1.253*
RWLE	300	0.067; 0.067*	0.147; 0.179*	0.083; 0.089*	0.147; 0.189*
	800	0.051; 0.054*	0.118; 0.149*	0.051; 0.058*	0.108; 0.152*

In the absence of outliers, in general, all estimators present similar behaviour. Relating to the estimation of the variance of the innovations, the MLE and WLE seem to be more precise which is an expected result since the data is Gaussian with zero-mean and these two methods are asymptotically equivalents. The RMSE of the estimators decreases as the sample size increases. When the simulated data has outliers, as an expected result the MLE and WLE estimates are totally corrupted by the atypical observations while the RWLE estimator presents generally accurate estimates. This simple example of simulation leads to the same conclusions of the models discussed previously in which $M-$regression method was also considered.

The methods discussed above give strong motivation to use the methodology in practical situations in which periodically correlated time series contain additive outliers. For example, Sarnaglia et al. (2010) applied the robust ACF estimator $\hat{\gamma}_Q^{(\nu)}(h)$ to fit a model for the quarterly Fraser River data. Sarnaglia et al. (2016) and Solci et al. (2018) analysed air pollution variables using the robust methodologies discussed in these papers. In the first paper, the authors considered the daily average SO_2 concentrations and, in the second one, it was analysed the daily average PM_{10} concentrations. Both data set were collected at Automatic Air Quality Monitoring Network (RAMQAr) in the Great Vitória Region GVR-ES, Brazil, which is composed by nine monitoring stations placed in strategic locations and accounts for the measuring of several atmospheric pollutants and meteorological variables in the area. In general, the models well fitted the series and all these applied examples revealed outliers effects on the estimates.

5 Proof of Theorem 1

By Propositions 1 and 4 and Example 1 of Wu (2007) the assumptions of Theorem 1 of Wu (2007) hold. Thus,

$$\sqrt{\frac{N}{2}}(F(c)-F(-c))\hat{\boldsymbol{\beta}}_N^{\mathrm{M}}(\lambda_j) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\boldsymbol{\Delta}^{(j)}\right),\ N\to\infty\ ,$$

with

$$\boldsymbol{\Delta}^{(j)}=\sum_{k\in\mathbb{Z}}\mathbb{E}\{\psi(\varepsilon_0)\psi(\varepsilon_k)\}\boldsymbol{\Delta}_k^{(j)},$$

where

$$\boldsymbol{\Delta}_k^{(j)}=\lim_{N\to\infty}\frac{2}{N}\sum_{\ell=1}^{N-|k|}\begin{pmatrix}\cos(\ell\lambda_j)\\ \sin(\ell\lambda_j)\end{pmatrix}\left(\cos((\ell+k)\lambda_j)\ \sin((\ell+k)\lambda_j)\right).$$

Observe that

$$\begin{aligned}\boldsymbol{\Delta}_k^{(j)}&=\lim_{N\to\infty}\frac{2}{N}\sum_{\ell=1}^{N-|k|}\begin{pmatrix}\frac{\cos(k\lambda_j)+\cos((2\ell+k)\lambda_j)}{2} & \frac{\sin(k\lambda_j)+\sin((2\ell+k)\lambda_j)}{2}\\ \frac{-\sin(k\lambda_j)+\sin((2\ell+k)\lambda_j)}{2} & \frac{\cos(k\lambda_j)-\cos((2\ell+k)\lambda_j)}{2}\end{pmatrix}\\ &=\begin{pmatrix}\cos(k\lambda_j) & \sin(k\lambda_j)\\ -\sin(k\lambda_j) & \cos(k\lambda_j)\end{pmatrix}+\lim_{N\to\infty}\frac{2}{N}\sum_{\ell=1}^{N-|k|}\begin{pmatrix}\frac{\cos((2\ell+k)\lambda_j)}{2} & \frac{\sin((2\ell+k)\lambda_j)}{2}\\ \frac{\sin((2\ell+k)\lambda_j)}{2} & \frac{-\cos((2\ell+k)\lambda_j)}{2}\end{pmatrix}.\end{aligned}$$

By observing that

$$\begin{aligned}&\frac{1}{N}\sum_{\ell=1}^{N-|k|}\cos((2\ell+k)\lambda_j)=\frac{\cos(k\lambda_j)}{N}\sum_{\ell=1}^{N-|k|}\cos(2\ell\lambda_j)+\frac{\sin(k\lambda_j)}{N}\sum_{\ell=1}^{N-|k|}\sin(2\ell\lambda_j)\\ &=\frac{\cos(k\lambda_j)}{N}\cos(\lambda_j(N-|k|-1))\frac{\sin(\lambda_j(N-|k|))}{\sin(\lambda_j)}+\frac{\sin(k\lambda_j)}{N}\sin(\lambda_j(N-|k|-1))\frac{\sin(\lambda_j(N-|k|))}{\sin(\lambda_j)}\end{aligned}$$

tends to zero as N tends to infinity and that the same holds for $N^{-1}\sum_{\ell=1}^{N-|k|}\sin(2\ell+k)$, this concludes the proof.

Acknowledgements

V. A. Reisen gratefully acknowledges partial financial support from FAPES/ES, CAPES/Brazil and CNPq/Brazil and CentraleSupélec. Màrton Ispàny was supported by the EFOP-3.6.1-16-2016-00022 project. The project is cofinanced by the European Union and the European Social Fund. Paulo Roberto Prezotti Filho and Higor Cotta are Ph.D students under supervision of V. A. Reisen and P. Bondon. The authors would like to thank the referee for the valuable suggestions.

Bibliography

Basawa, I., Lund, R., 2001. Large sample properties of parameter estimates for periodic ARMA models. Journal of Time Series Analysis 22 (6), 651–663.

Croux, C., Rousseeuw, P. J., 1992. Time-efficient algorithms for two highly robust estimators of scale. Computational Statistics 1, 1–18.

Fajardo, F., Reisen, V. A., Cribari-Neto, F., 2009. Robust estimation in long-memory processes under additive outliers. Journal of Statistical Planning and Inference 139 (8), 2511–2525.

Fajardo, F. A., Reisen, V. A., Lévy-Leduc, C., Taqqu, M., 2018. M-periodogram for the analysis of long-range-dependent time series. Statistics, 665–683.

Fox, A. J., 1972. Outliers in time series. Journal of the Royal Statistical Society 34 (B), 350–363.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series model. Journal of Time Series Analysis 4 (4), 221–238.

Huber, P., 1964. Robust estimation of a location parameter. The Annals of Mathematical Statistics 35, 73–101.

Koul, H. L., Surgailis, D., 2000. Second order behavior of M-estimators in linear regression with long-memory errors. Journal of Statistical Planning and Inference 91 (2), 399–412.

Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M., Reisen, V. A., 2011. Robust estimation of the scale and the autocovariance function of Gaussian short and long-range dependent processes. Journal of Time Series Analysis 32 (2), 135–156.

Ma, Y., Genton, M., 2000. Highly robust estimation of the autocovariance function. Journal of Time Series Analysis 21 (6), 663–684.

McLeod, A. I., 1994. Diagnostic checking periodic autoregression models with application. Journal of Time Series Analysis 15 (2), 221–33.

Reisen, V. A., Lévy-Leduc, C., Taqqu, M. S., 2017. An M-estimator for the long-memory parameter. Journal of Statistical Planning and Inference 187, 44 – 55.

Rousseeuw, P. J., Croux, C., 1993. Alternatives to the median absolute deviation. Journal of the American Statistical Association 88 (424), 1273–1283.

Sarnaglia, A. J. Q., Reisen, V. A., Bondon, P., 2015. Periodic ARMA models: Application to particulate matter concentrations. In: 23rd European Signal Processing Conference. pp. 2181–2185.

Sarnaglia, A. J. Q., Reisen, V. A., Bondon, P., Lévy-Leduc, C., 2016. A robust estimation approach for fitting a PARMA model to real data. In: IEEE Statistical Signal Processing Workshop. pp. 1–5.

Sarnaglia, A. J. Q., Reisen, V. A., Lévy-Leduc, C., 2010. Robust estimation of periodic autoregressive processes in the presence of additive outliers. Journal of Multivariate Analysis 101 (9), 2168–2183.

Shao, Q., 2008. Robust estimation for periodic autoregressive time series. Journal of Time Series Analysis 29 (2), 251–263.

Solci, C. C., Reisen, V. A., Sarnaglia, A. J. Q., Pascal, B., 2018. Empirical study of robust estimation methods for PAR models with application to PM_{10} data. in press, Communication in Statistics 15.

Wu, W. B., 2007. M-estimation of linear models with dependent errors. The Annals of Statistics 35 (2), 495–521.