Vol. 00, No. 00, Month 20XX, 1–21

M-periodogram for the analysis of long-range dependent time series

F.A. Fajardo\textsuperscript{a}, V. A. Reisen\textsuperscript{b,e}, C. Lévy-Leduc\textsuperscript{c} and M.S. Taqqu\textsuperscript{d}

\textsuperscript{a}DEST-Universidade Federal do Espirito Santo, Brazil; \textsuperscript{b}DEST and PPGEA-Universidade Federal do Espirito Santo, Brazil; \textsuperscript{c}UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France; \textsuperscript{d}Department of Mathematics, Boston University; \textsuperscript{e}Laboratoire des Signaux et Systèmes, CNRS, Université Paris-Saclay, France.

(Received 00 Month 20XX; final version received 00 Month 20XX)

This paper focuses on time series with long-memory and suggests using an alternative periodogram, called $M$-periodogram, which is obtained by relating the periodogram to a regression problem and then using an $M$-estimator for the coefficients of the regression model. The asymptotic properties of this novel $M$-periodogram are established and its empirical properties are investigated for finite samples under different scenarios. Furthermore, in addition to being an appealing alternative periodogram for long-memory time series, it is also resistant to additive outliers. We investigate the robustness performance of the estimator through simulation. As a practical application, the paper investigates the effect of atypical observations in air pollution data, namely, daily Particulate Matter ($PM_{10}$) observations. Besides the importance of modeling and forecasting this pollutant, the $PM_{10}$ series presents, in general, interesting features such as seasonal poles, asymmetry, and also high levels of pollution which can be regarded as atypical observations in the context of this work.

Keywords: Time Series, robust estimation, long memory, outliers, pollution.

1. Introduction

The classical periodogram is a powerful tool in the spectral analysis of time series. It is well-known, however, that it is very sensitive to the presence of outliers (Fox (1972)), and therefore, it becomes useless in situations where the real data is contaminated by atypical observations. Since additive outliers are quite common in practice, defining a new periodogram which is not sensitive to the presence of additive outliers has a real practical interest.

Several approaches have been proposed in the time series literature in order to deal with outliers. Under a Gaussian assumption for the core process, that is, the uncontaminated time series, Tatum and Hurvich (1993a) and Tatum and Hurvich (1993b) proposed interesting ways to construct high breakdown methods for handling additive contamination based on robust trigonometric regression to obtain a robustified discrete Fourier transform. Their approaches consist in obtaining a filtered version of the data to reconstruct the “core process”, namely the original underlying process without the outliers. References related to computing robust spectral estimators include, for example, Katkovnik (1998), Li (2008) and references therein. More generally, Kleiner et al. (1979), Martin and Thomson (1982) and more recently Spangl and Dutter (2013), Hagemann

*Corresponding author. Email: celine.levy-leduc@agroparistech.fr
provide several approaches for dealing with robust spectral analysis. Most of these references, however, concern weakly dependent time series.

In the long-memory framework, Fajardo et al. (2009) suggested a robust plug-in periodogram, that is, a periodogram obtained by replacing the standard sample autocovariance by the robust autocovariance given in Ma and Genton (2000). In order to deal with outliers, Fajardo et al. (2009) introduced a robust estimator of the memory parameter $d$ in the ARFIMA model which is defined in (25). That paper motivated the research of Sarnaglia et al. (2010), Lévy-Leduc et al. (2011), Lévy-Leduc et al. (2011), Lévy-Leduc et al. (2011), Reisen and Fajardo (2012) among others. Sarnaglia et al. (2010) introduced a robust method to estimate, through the Yule-Walker equations, the parameters of periodic autoregressive (PAR) models. Asymptotic properties of the robust autocovariance function, for short and long-range dependence, motivated the works Lévy-Leduc et al. (2011), Lévy-Leduc et al. (2011), Lévy-Leduc et al. (2011). A review of robust methods in the frequency domain and their use in applied works can be found in Reisen and Fajardo (2012).

To assess the accuracy of the asymptotic results on the robust estimation of the autocovariance and on the parameters of the models, simulation studies for finite sample sizes, in different scenarios, were discussed in the previous references. In general, the empirical studies show that the estimators, based on the robust autocovariance, are outlier resistant and very useful in applied works. However, the robustness properties of the estimators weaken when the data deviates from the Gaussian assumption, for example, when the series has a heavy-tailed distribution. This may be explained by the fact that the robust autocovariance function involves a constant which depends strongly on the Gaussian distribution assumption. In addition, the robust autocovariance estimator does not have the non-negative definite property, and becomes useless in the context of a non-stationary process such as, for example, the ARFIMA process with $0.5 < d < 1.0$. This parameter range is quite often encountered in time series with long-memory (Hurvich and Ray (1995), Velasco (1999), Franco and Reisen (2007) among others).

In this paper, we propose an alternative to the robust periodogram considered in Fajardo et al. (2009) for long-memory processes. More precisely, this paper suggests relating the periodogram to a regression problem and the new periodogram, here denoted $M$-periodogram, is built by using $M$-estimators, thus extending the approach of Li (2010). Li suggests the use of the $L^p$-norm periodogram for some $p > 1$ and $p \neq 2$ for weakly dependent times series. Here, the $M$-periodogram deals with long-range dependent time series and is not based on the autocorrelation. Note that, the method proposed here computes the periodogram directly from the data using the $M$-regression approach. This spectral estimator has recently been widely used in different contexts of short-memory time series modelling. Therefore, the theoretical and empirical results discussed in this paper contribute to fill the gap in long-memory processes.

The outline of the paper is as follows. In Section 2 some theoretical properties of the new $M$-periodogram are established based on the results of Koul (1992), who was a pioneer in establishing large sample properties of $M$-estimators in linear models with long-range dependent errors. Finite sample size investigation is addressed in Section 3. An illustrative example is presented in Section 4 and the proofs are given in Section 6.
2. Theoretical results

For a given time series \( Y_1, \ldots, Y_N \), the classical periodogram at Fourier frequencies \( \lambda_j = 2\pi j/N, \ j = 1, \ldots, [N/2] \), is defined by

\[
I_N(\lambda_j) = \frac{1}{2\pi N} \left| \sum_{k=1}^{N} Y_k \exp(ik\lambda_j) \right|^2.
\]

First, observe that \( I_N(\lambda_j) \) is related to the least-square estimate \( \hat{\beta}_{N}^{LS}(\lambda_j) \) of a two-dimensional vector \( \beta \) in the linear regression model

\[
Y_i = c^T_N i \beta + \varepsilon_i, \ 1 \leq i \leq N, \ \beta \in \mathbb{R}^2,
\]

where

\[
c^T_N i (\lambda_j) = (\cos(i\lambda_j) \sin(i\lambda_j)) ,
\]

and where \( \varepsilon_i \) denotes the deviation of \( Y_i \) from \( c^T_N i \beta \).

Then

\[
\hat{\beta}_{N}^{LS}(\lambda_j) = \text{Arg min}_{\beta \in \mathbb{R}^2} \sum_{i=1}^{N} (Y_i - c^T_N i (\lambda_j) \beta)^2 .
\]

Indeed, the solution of (3) is

\[
\hat{\beta}_{N}^{LS}(\lambda_j) = (C^T C)^{-1} C^T Y = \frac{2}{N} C^T Y = \frac{2}{N} \left( \sum_{k=1}^{N} Y_k \cos(k\lambda_j) \sum_{k=1}^{N} Y_k \sin(k\lambda_j) \right)^T,
\]

where \( Y = (Y_1, \ldots, Y_N)^T \) and where \( C \) and \( C^T C \) are defined by

\[
C = \begin{pmatrix}
\cos(\lambda_j) & \sin(\lambda_j) \\
\cos(2\lambda_j) & \sin(2\lambda_j) \\
\vdots & \vdots \\
\cos(N\lambda_j) & \sin(N\lambda_j)
\end{pmatrix}
\]

and

\[
C^T C = \left( \sum_{k=1}^{N} \cos(k\lambda_j)^2 \sum_{k=1}^{N} \cos(k\lambda_j) \sin(k\lambda_j) \right) = \frac{N}{2} \text{Id}_2,
\]

respectively, with \( \text{Id}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \).

Hence

\[
I_N(\lambda_j) = \frac{N}{8\pi} \| \hat{\beta}_{N}^{LS}(\lambda_j) \|^2 =: I_{N}^{LS}(\lambda_j),
\]

where \( \| \cdot \| \) denotes the classical Euclidian norm. It can thus be seen from (7) that there is a connection between the classical periodogram and the estimator of \( \beta \) in the linear
regression model \([1]\). This observation underlies our methodology. To obtain an \(M\)-periodogram, an \(M\)-estimator of \(\beta\) will be used.

Note that the connection (7) is valid for any choice of \(\varepsilon_i, i = 1, \ldots, N\). We will suppose here that

\[
\varepsilon_i = G(\eta_i) ,
\]

with \(E(\varepsilon_i) = 0\) and \(E(\varepsilon_i^2) < \infty\). In (8), \(G\) is a non null-valued and skew symmetric measurable function (i.e. \(G(-x) = -G(x)\), for all \(x\)) and \((\eta_i)_{i \geq 1}\) is a long-range dependent stationary zero-mean Gaussian process with unit variance satisfying the following assumption (A1).

(A1) \((\eta_i)_{i \geq 1}\) is a stationary mean-zero Gaussian process with covariances \(\rho(k) = E(\eta_i \eta_{k+1})\) satisfying:

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

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:

\[
f(\lambda) = |1 - \exp(-i\lambda)|^{-2d} f^*(\lambda) ,
\]

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

\[
D = 1 - 2d ,
\]

where \(D\) is defined in Assumption (A1). The fact that \((\eta_i)_{i \geq 1}\) is required to satisfy (A1) essentially means that both \(L(x), x \geq 1\) and \(f^*(\lambda), \lambda \in (-\pi, \pi]\) satisfy some smoothness properties.

In the following, an \(M\)-periodogram based on an \(M\)-estimator \(\hat{\beta}_N\) of the regression coefficient \(\beta\) in \([1]\) is built. The \(M\)-estimator \(\hat{\beta}_N\) is defined as the solution \(t\) of

\[
\sum_{i=1}^N c_{Ni} \psi(Y_i - c_{Ni}^T t) = 0 ,
\]

where, as in [Koul (1992)], \(\psi\) is a real-valued measurable function satisfying

(A2) \(0 < E[\psi^2(\varepsilon_1)] < \infty\).

(A3) The function \(\psi\) is absolutely continuous with its almost everywhere derivative \(\psi'\) satisfying \(E[|\psi'(\varepsilon_1)|] < \infty\) and such that the function \(z \mapsto E[|\psi'(\varepsilon_1 - z) - \psi'(\varepsilon_1)|]\) is continuous at zero.

(A4) \(\psi\) is nondecreasing, \(E[\psi'(\varepsilon_1)] > 0\) and \(E[\psi'(\varepsilon_1)^2] < \infty\).

(A5) \(\psi\) is skew symmetric, i.e. \(\psi(-x) = -\psi(x)\), for all \(x\).
Observe that, since $G$ in (8) is assumed to be skew symmetric, (A5) implies that $\mathbb{E}[\psi(\varepsilon_1)] = 0$.

Consider the particular case where $Y_i$ is defined in (1) with $\beta = 0$. Hence,

$$Y_i = \varepsilon_i$$

with

$$\varepsilon_i = G(\eta_i)$$

defined in (8).

The following theorem establishes the asymptotic properties of the corresponding $M$-estimator $\hat{\beta}_N(\lambda_j)$.

**Theorem 2.1** Assume that (A1), (A2), (A3), (A4) and (A5) hold and that $\beta = 0$ in (1) so that $Y_i = \varepsilon_i$. Then, for any fixed $j$, $\hat{\beta}_N(\lambda_j)$ defined by (11) satisfies

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

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

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

where

$$\tilde{\Gamma} = \lim_{N \to \infty} \frac{4}{N^{2-D}} \sum_{1 \leq k, \ell \leq N} c_{Nk}(\lambda_j)c_{N\ell}(\lambda_j)\rho(k - \ell)$$

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

In Relation (14), the vector $c_{N}(\lambda_j)$ is defined in (2).

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

and

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

The proof of Theorem 2.1 is given in Section 6.1. This theorem has a number of important implications which are given in the following corollaries.

**Remark 2.2** Observe that the standard least-squares estimator $\hat{\beta}_N^{LS}(\lambda_j)$ is a particular $M$-estimator satisfying (11) when $\psi(x) = x$ for all $x$ in $\mathbb{R}$. Indeed, another way of
expressing \((3)\) is to require, after taking a derivative with respect to \(\beta\) and setting it equal to 0, that

\[
\sum_{i=1}^{N} c_{Ni}(Y_i - c_{Ni}^T \beta) = 0,
\]

which is \((11)\) with \(\psi(x) = x\) and \(t = \beta\).

**Remark 2.3** Theorem 2.1 is established under the assumption that the errors \((\epsilon_i)\) are subordinated to a long memory Gaussian process. However, following Giraitis et al., 2012, Theorem 11.2.2, Theorem 2.1 would also be valid if the \((\epsilon_i)\) were assumed to be long memory moving average errors of the following form: \(\epsilon_j = \sum_{k \geq 0} a_k \zeta_{j-k}\), where the \((\zeta_t)\) are assumed to be i.i.d. with zero mean and finite variance.

**Remark 2.4** Note that in Theorem 2.1 \(j\) is assumed to be fixed as \(N \to \infty\) since the main important frequencies for long-memory processes are those close to zero, that is, the long memory property of the process is more pronounced at low frequencies. This issue will be further discussed in the Application Section 4.

**Corollary 2.5** Under the assumptions of Theorem 2.1 and when \(G(x) = x \in \mathbb{R}\), that is, in the Gaussian case \(G(\eta) = \eta\), the ratio \(J_1^2 / (E[\psi'(\epsilon_1)]^2)\) appearing in \((13)\) equals 1. Moreover, \(\hat{\beta}_N(\lambda_j)\) defined by \((14)\) has the same asymptotic behavior as \(\hat{\beta}_N^{LS}(\lambda_j)\), namely \(\hat{\beta}_N\) is asymptotically first order equivalent to the least squares estimator \(\hat{\beta}_N^{LS}\).

The proof of Corollary 2.5 is given in Section 6.2

Following Li (2008, 2010), the \(M\)-periodogram here proposed is defined as

\[
I_{N,\psi}(\lambda_j) = \frac{N}{8\pi} \| \hat{\beta}_N(\lambda_j) \|^2,
\]

where \(\|x\|\) denotes the standard Euclidean norm. If we let \(\psi(x) = x\), then Equation \((18)\) becomes the classical periodogram.

**Corollary 2.6** Under the assumptions of Theorem 2.1, the periodogram \(I_{N,\psi}\) defined in \((18)\) satisfies

\[
N^{D-1} I_{N,\psi}(\lambda_j) \rightarrow_d (Z_1^2 + Z_2^2), \text{ as } N \to \infty,
\]

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

\[
\frac{J_1^2}{8\pi (E[\psi'(\epsilon_1)]^2)^2} \tilde{\Gamma},
\]

with \(\tilde{\Gamma}\) defined in \((14)\).

The proof of Corollary 2.6 is given in Section 6.2
Corollary 2.7  Under the assumptions of Theorem 2.1 as \( N \rightarrow \infty \),

\[
\frac{I_{N,\psi}(\lambda_j)}{f(\lambda_j)} \xrightarrow{d} \frac{J_1^2}{(\mathbb{E}[\psi'(\varepsilon_1)])^2} (\mathcal{L}_1 W_1^2 + \mathcal{L}_2 W_2^2),
\]

(21)

where \( W_1 \) and \( W_2 \) are zero-mean independent Gaussian random variables and \( \mathcal{L}_1, \mathcal{L}_2 \) are defined in (16) and (17), respectively. In addition,

\[
\log \left( \frac{I_{N,\psi}(\lambda_j)}{f(\lambda_j)} \right) \xrightarrow{d} \log \left( \frac{J_1^2}{(\mathbb{E}[\psi'(\varepsilon_1)])^2} \right) + \log \left( \mathcal{L}_1 W_1^2 + \mathcal{L}_2 W_2^2 \right).
\]

(22)

The proof of Corollary 2.7 is given in Section 6.2.

Remark 2.8 Suppose that the assumptions of Theorem 2.1 hold and that \( G(x) = x \) in (8), that is, in the Gaussian case, it then follows from Corollary 2.5 that \( I_{N,\psi}(\lambda_j) \) and \( I_{N,\psi}(\lambda_j) \) have the same asymptotic distribution. See, also Theorem 6 of Hurvich and Beltrão (1993).

Remark 2.9 We assume here that the process is long-range dependent. For asymptotic results for the \( M \)-periodogram in the short-range dependent case see, for example, Li (2010).

Typical functions \( \psi \) include the Huber function defined by

\[
\psi(x) = \begin{cases} 
  x & \text{if } |x| \leq k, \\
  \text{sign}(x)k & \text{if } |x| > k,
\end{cases}
\]

(23)

and the \( L_p \) function

\[
\psi(x) = p|x|^{p-1}\text{sign}(x), \text{ for } 1 < p \leq 2.
\]

(24)

The choice (23) is referred here as the “Huber periodogram”, since the function \( \psi \) defined in (23) corresponds to the classical Huber function. The choice (24) is denoted here by “\( L_p \) periodogram”, since (24) corresponds to the first derivative of the \( L_p \) norm to the power \( p \). When \( p = 2 \), the associated \( \hat{\beta}_N \) corresponds to the estimator obtained by the usual least-square approach leading to the classical periodogram. Note that the \( L_p \) periodogram has been proposed by Li (2008, 2010) for \( p \) between 1 and 2. In these papers, the author shows that the \( L_p \) periodogram is a robust estimator of the spectral density of weakly dependent processes for dealing with heavy-tailed observations. This is the reason why the comparison with the \( L_p \) periodogram is included in Section 3 even if the \( L_p \) function is not bounded.

Theorem 2.10 If the c.d.f \( F_{\varepsilon} \) of \( \varepsilon_1 \) defined in (8) is a continuous and increasing function, then the results of Theorem 2.1 hold for the Huber function \( \psi \) defined in (23).

The proof of Theorem 2.10 is given in Section 6.3.

Remark 2.11 Note that in the case of the \( L_p \) periodogram for \( p \in (1, 2) \), Assumption (A3), which is required for Theorem 2.1 may not be satisfied.
3. Numerical experiments

Let \( \{x_t\}_{t=1,...,N} \) be a sample from a Gaussian stationary ARFIMA(0,d,0) process \( (X_t) \) with innovations with zero mean and variance one, that is, \( (X_t) \) is defined by

\[
X_t = (1 - B)^{-d}Z_t = \sum_{j \geq 0} \frac{\Gamma(j + d)}{\Gamma(j + 1) \Gamma(d)} Z_{t-j},
\]

where \( (Z_t) \) are i.i.d. standard Gaussian random variables.

Let \( \{y_t\} \) be a sample of the process contaminated by additive outliers, defined by

\[
y_t = x_t + \omega W_t
\]

where the parameter \( \omega \) 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 \( \mathbb{E}[W_t^2] = \text{Var}(W_t) = \delta \). Note that (26) is based on the parametric models proposed by Fox (1972). \( W_t \) is the product of Bernoulli(\( \delta \)) and Rademacher random variables; the latter equals 1 or \(-1\), both with probability \( \frac{1}{2} \).

The samples \( \{x_t\} \) were generated with \( d = 0.1 \) and 0.45, \( N = 100 \) and 500 and the contaminated data \( y_t \) were generated from (26) with \( \delta = 0.05, 0.10 \) for magnitudes \( \omega = 0 \) (no outliers) and 10. Other values of \( d \) between 0.1 and 0.45 and \( \omega = 3, 5 \) were also considered. Since the performance of the method were similar to the above cases they are not presented here but are available upon request. The periodograms \( I_{N,\psi} \) defined in (18) were computed for the Huber function \( \psi(.) \) defined in (23) with \( k = 1.345 \) (see the example and discussion in Sections 7.5 (page 378) and 8.2, respectively, of Hampel et al. (1986)) and \( p = 1.2, 1.7, 2.0 \) for the \( L_p \) functions, defined in (24). The Laplace periodogram proposed by Li (2008) is also considered for comparison purpose since this method handles heavy-tailed noise and non-linear distortion and since it has not been empirically studied yet in the context of long-memory processes with additive outliers.

In addition, in practical situations, it is more convenient to consider \( p = 1 \) instead of \( 1 < p < 2 \). In that case, \( \psi(x) = \text{sign}(x) \). Therefore, it becomes interesting to also verify the empirical performance of this method under the scenario of long-memory process with additive outliers.

In the tables and figures, the periodograms have the following notations: HU for the Huber function \( \psi(.) \) defined in (23), \( L_p \) for the \( L_p \) function \( \psi(.) \) defined in (24) and LAD for the Laplace periodogram which is defined thanks to the function \( \psi \) of (24) with \( p = 1 \). For further details, we refer the reader to Li (2008). As previously mentioned, the case \( p = 2 \) corresponds to the ordinary periodogram. The results in figures (the empirical mean) and tables (the root of mean square error values (RMSE)) are based on 1000 replications and all simulations were carried out using the Ox matrix programming language (see http://www.doornik.com). The empirical RMSE corresponds to the mean over all values of the mean of \( \text{RMSE}_j \) for the Fourier frequencies at \( j = 1, ..., \lfloor \frac{N}{2} \rfloor - 1 \).
\[ RMSE = \frac{1}{m} \sum_{j=1}^{m} RMSE_j \]  

(27)

where

\[ RMSE_j = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left( I_{N,\psi}^{(k)}(\lambda_j) - f(\lambda_j) \right)^2}, \]

where \( I_{N,\psi}^{(k)}(\lambda_j) \) corresponds to the \( k \)th replication \((k = 1, \ldots, K)\), \( f(.) \) is the spectral density of the ARFIMA(0,\( d \),0) process, \( K = 1000 \) and \( m = \lceil N/2 - 1 \rceil \).

The empirical investigation is divided into two parts. In the first part, the asymptotic results discussed in the previous section are empirically investigated, that is, the behavior of the \( M \)-periodograms in the context of non-contaminated data. In the second part, the purpose is to evaluate the robustness properties of the estimators under different scenarios of series corrupted by outliers. Figure 1 displays a single realization of a contaminated ARFIMA(0,\( d \),0) process with \( \delta = 0.05 \), \( \omega = 10 \) and \( N = 100 \).

![Figure 1](image)

Figure 1. A contaminated ARFIMA(0,0.1,0) series with \( \delta = 0.05 \), \( \omega = 10 \) and \( N = 100 \).

Table 1 displays the RMSE values defined by (27) in the case of non-contaminated data. The entries in this table show that, even for a small sample size and \( d \) close to the non-stationary boundary \((d = 0.45)\), the \( M \)-periodograms (HU, \( L_p \), \( p = 1.2, 1.7 \)) fare reasonably well relative to the ordinary periodogram \((L_2)\). As the sample size increases, all periodograms presented similar RMSE values, as would be expected from the asymptotic results. An \( L_p \) periodogram with \( p = 1.7 \) delivers more accurate estimates than for \( p = 1.2 \), which is expected since it is closer to the ordinary periodogram, which seems to be an alternative choice when the data is not contaminated. Surprisingly, the performance of the Laplace periodogram (LAD) are not on a par with those of the other robust periodograms except in the case \( d = 0.1 \) for which they are close to those of the \( L_{1.2} \)-periodogram.

The \( M \)-periodogram is at its greatest advantage, obviously, when the data departs from the standard assumption, that is, when the data has outliers and/or heavy-tailed distribution. The discussion of the robustness property of the proposed estimator, in series under outlier contamination, is based on the empirical results that are given in
Tables 2 and 3 and Figures 2 and 3. As expected, the ordinary periodogram ($L_2$) is totally corrupted by the outliers and, therefore, its use should be avoided when the series contains additive outliers. On the other hand, the $M$ estimators present generally accurate estimates even for a large number of outliers. In contrast to the performance of the robust method under non-contaminated series, the most accurate robust method is attained for the Huber periodogram and secondly by the $L_p$ periodogram with $p = 1.2$ in the case where $d = 0.1$. This was also noted by [Li (2010)] in the short-range dependence case.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\delta$</th>
<th>HU</th>
<th>LAD</th>
<th>$L_{1.2}$</th>
<th>$L_{1.7}$</th>
<th>$L_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.05</td>
<td>0.0437</td>
<td>0.1208</td>
<td>0.0940</td>
<td>0.3816</td>
<td>0.8238</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>0.0409</td>
<td>0.1059</td>
<td>0.0589</td>
<td>0.3101</td>
<td>0.7838</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
<td>0.0502</td>
<td>0.2124</td>
<td>0.3260</td>
<td>0.7031</td>
<td>1.2405</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>0.0761</td>
<td>0.1485</td>
<td>0.0848</td>
<td>0.5993</td>
<td>1.6117</td>
</tr>
</tbody>
</table>

Table 2. The empirical $RMSE$ of the $M$-periodogram estimators defined by (27) for the contaminated ARFIMA(0,0.1,0) process.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\delta$</th>
<th>HU</th>
<th>LAD</th>
<th>$L_{1.2}$</th>
<th>$L_{1.7}$</th>
<th>$L_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.05</td>
<td>0.1311</td>
<td>1.5745</td>
<td>0.8391</td>
<td>0.5026</td>
<td>0.7615</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>0.1926</td>
<td>0.5947</td>
<td>0.3863</td>
<td>0.4806</td>
<td>0.77989</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
<td>0.2556</td>
<td>2.2718</td>
<td>1.8356</td>
<td>2.1955</td>
<td>1.5119</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>0.2517</td>
<td>1.1965</td>
<td>0.9209</td>
<td>1.0583</td>
<td>1.6207</td>
</tr>
</tbody>
</table>

Table 3. The empirical $RMSE$ of the $M$-periodogram estimators defined by (27) for the contaminated ARFIMA(0,0.45,0) process.

<table>
<thead>
<tr>
<th>$d$</th>
<th>$L_2$</th>
<th>HU</th>
<th>LAD</th>
<th>$L_{2,c}$</th>
<th>HU,c</th>
<th>LAD,c</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.01464</td>
<td>0.02013</td>
<td>0.10132</td>
<td>0.89032</td>
<td>0.05183</td>
<td>0.12278</td>
</tr>
<tr>
<td>0.3</td>
<td>0.02248</td>
<td>0.02445</td>
<td>0.13111</td>
<td>0.77118</td>
<td>0.05398</td>
<td>0.15972</td>
</tr>
<tr>
<td>0.4</td>
<td>0.02363</td>
<td>0.03081</td>
<td>0.21980</td>
<td>0.97293</td>
<td>0.09645</td>
<td>0.44426</td>
</tr>
</tbody>
</table>

Table 4. The empirical $RMSE$ of the $M$-periodogram estimators for the uncontaminated and contaminated ARFIMA(0,d,0), $d = 0.2, 0.3, 0.4$, process when $N = 100$ and $\delta = 0.05$. The results associated with the contaminated case are displayed in the columns $\ast_c$.

For further comparison, Table 4 displays the RMSE values for contaminated and uncontaminated data for other values of $d$. 


One way to verify, in practical applications, whether the series has or has not departed from the assumption of no contamination is to look at the plot of the periodogram.

From Figures 2 and 3, one observes that the level (empirical mean) of the classical periodogram across the frequencies is much higher than the robust ones. Since this paper deals with long-memory series, it is interesting to examine also the behavior of the classical periodogram near zero frequency: note that its peaks are much larger than the robust ones. This phenomenon, expected from Corollary 1 of Fajardo et al. (2009), will also be observed and further discussed in the Application Section.

4. Application to pollution data

The spectral estimators discussed in the previous section are used here as tools to analyze the daily Particulate Matter (PM10) series (PM with an aerodynamic diameter of less than or equal to 10 µm, expressed in µg/m3, measured at the Automatic Air Quality
Monitoring Network (AAQMN) in the Greater Vitória Region (GRV), ES, Brazil. The AAQMN has eight monitoring sites located at Laranjeira, Cariacica, Vila Velha, Ibes, Sua and Vitória center. The choice of this data set was motivated by the fact that, besides the importance of modeling and forecasting this pollutant in the context of air quality control, the $PM_{10}$ series presents, in general, some interesting features. These will be investigated as a time series, analyzing stationarity, periodicity, long-memory at zero and seasonal poles, asymmetry and in particular considering possible atypical observations which would call for an application of the statistical methodology suggested here. The recent works by Reisen et al. (2014), Reisen et al. (2014) and Reisen et al. (2017) deal with this series to fit models under different context. The main interest here is to show the usefulness of the $M$-periodograms in a real-life application when dealing with a data set that does not meet the standard time series model assumptions, that is, a data set in which there are outliers. These outliers may be associated with observations that present high pollution levels, which is quite common in $PM_{10}$ series due too many factors, such as season or day of the week.

The raw series consist of daily mean of observations from March, 1st 2008 to December, 31st 2009 ($N = 671$). Figure 4 displays the $PM_{10}$ series obtained at the location Vila Velha. The dominant feature of the graph is that the series contains persistent oscillation with period of approximately 7 days, but with irregular variations in amplitude.

Figures 6 (a) and (b) display the plot of the classical $L^2$ and Huber periodograms, respectively, using the original data set. In general, the spectral estimators present large peaks concentrated at near zero frequency and also at frequencies which are integral multiples of $\frac{1}{7}$ (see the indication with (‘o’) in the Figure 6(a)). These plots also indicate that, among other properties such as the seasonality of period 7, the series appears to possess long memory. The period 7 is due the fact that the series corresponds to the daily mean observations. The long-memory property at zero frequency can also be seen in Figure 5, which displays the log periodogram versus the log frequency (bandwidth of $N^{0.65}$), where the points are scattered around the negative slope with estimated value of -0.631 and standard deviation (sd) equals to 0.172. This is typically the case for long-memory processes. This bandwidth was chosen to avoid the seasonal effect. One interesting distinction between the plot of the standard periodogram in Figure 6(a) and the robust ones is the vertical scale difference. The estimate of the slope in the periodogram has smaller intensity in terms of absolute value than the robust one, which is -0.798 (sd= 0.163) (see Figure 7). It is important to ask how representative these differences are in a real application since the original series may contain outliers. As pointed out and theoretically justified in Proposition 2 and Corollary 1 in Fajardo et al. (2009) and also discussed in Reisen et al. (2017), the periodogram of a time series with additive outliers displays higher values than the one computed from the uncontaminated process.

In addition, to verify the effect of the high contaminant levels of a $PM_{10}$ series, that is, if the high levels of the contaminant can be or not identified as outliers, a new series was generated from the $PM_{10}$ data of the Vila Velha site as follows: The suspected atypical observations, that is, those observations bigger than $\mu + 3s$, where $\mu$ and $s$ denote the empirical mean and the empirical standard deviation of the original series respectively, were suppressed and replaced by the mean of the original series. The periodogram was computed for this new series and the plot is in Figure 6(c). In this context, it is clearly visible that the $L^2$-periodogram displays a similar magnitude as the Huber periodogram (Figure 6(b)). In addition, the estimates of the slopes are very close, they are -0.798 (sd =0.163) and -0.725 (sd = 0.145) for the $M$ and classical periodogram, respectively (see Figures 7 and 8). This corroborates the above discussion and suggests that the replaced
observations in the original series can be seen as outliers.

For further comparison, the $L_2$, Huber and LAD periodograms are applied to the Laranjeiras station data. The results are displayed in Figure 9. It is observed from this figure that the Huber and LAD periodograms seem to have the same behavior.

5. Conclusion

This paper proposes the $M$-periodogram estimator to estimate the spectral density of a time series with long-memory property. The assumption on the errors ($\varepsilon_i$) is not restricted to a Gaussian process, that is, it may be a function of a zero-mean Gaussian process ($\eta_i$) with long-memory property. Asymptotic properties of the proposed spectral estimators are established and simulations are provided to show the performance of the estimator under different scenarios. The robustness of the method is also investigated in the simulation study which also considers other alternative spectral estimators. The $M$-periodogram appeared to be better than the other spectral estimators studied, especially when the series is contaminated with outliers. Therefore, the proposed method becomes a good alternative to estimate the spectral density of a long-memory process and, consequently, its long-memory parameter $d$, as explored in Reisen et al. (2017). A data set related to air pollution variable is used to show the usefulness of the proposed methodology in real applications.

As pointed out by one referee, there are several robust regression methods apart from those here discussed that can be used as alternative approaches to develop robust periodogram estimators, one idea is to follow the Biweight transform and filter method discussed in Tatum and Hurvich (1993a, 1993b) and mentioned in the introduction. This among other methods discussed in Tatum and Hurvich (1993a) and Tatum and Hurvich (1993b) are future projects.
Figure 6. (a) Standard ($L_2$) periodogram plot of $PM_{10}$ concentrations at Vila Velha station, (b) Huber periodogram plot for $PM_{10}$ concentrations at Vila Velha station, (c) Standard ($L_2$) periodogram plot of the new $PM_{10}$ concentrations at Vila Velha station.

6. Proofs

6.1. Proof of Theorem 2.1

The proof of Theorem 2.1 involves the expansion of the function $\psi \circ G$ in (8) in Hermite polynomials. The first few are $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, .... For the convenience of the reader we shall first recall Theorem 3.2 of Koul (1992) where we kept the original labeling for easy reference. The theorem is formulated as:

**Theorem 6.1** Let $(Y_i)_{1 \leq i \leq N}$ satisfy (1) where $(\varepsilon_i)_{1 \leq i \leq N}$ are defined by (8) and (A1). Assume also that the following conditions hold:

(L1) $(Q^T Q)^{-1}$ exists for all $N \geq 2$, where $Q$ is the $N \times 2$ matrix having $q_{iN_1}^T = c_{N_1}^T$ as $i$th row, where $c_{N_1}^T$ is defined in (7).
Figure 7. Log $M$-periodogram plots for $PM_{10}$ concentrations at Vila Velha station.

Figure 8. Log-periodogram plots for the modified data of $PM_{10}$ concentrations at Vila Velha station.

Figure 9. Top left: Standard ($L_2$) periodogram plot of $PM_{10}$ concentrations at Laranjeiras station, top right: Standard ($L_2$) periodogram plot of the new $PM_{10}$ concentrations at Laranjeiras station, bottom left: Huber periodogram plot for $PM_{10}$ concentrations at Laranjeiras station, bottom right: LAD periodogram plot for $PM_{10}$ concentrations at Laranjeiras station.

$$(\pi 1) \limsup_{N \to \infty} \{ N \max \| d_i \|^2 \} < \infty, \text{ where } d_i = D_N^{-1} q_{Ni}, \text{ with } D_N = \text{diag}(\| q_1 \|, \ldots, \| q_p \|), q_j \text{ denoting the } j\text{th column of } Q.$$

Let us also consider a function $\psi$ satisfying the following conditions: $\psi$ is a measurable function from $\mathbb{R}$ to $\mathbb{R}$ such that:

$$(1.3) \mathbb{E}(\psi(\varepsilon_1)) = 0 \text{ and } 0 < \mathbb{E}(\psi^2(\varepsilon_1)) < \infty.$$

$$(L2) \psi \text{ is an absolutely continuous functions such that its almost everywhere derivative } \psi' \text{ satisfies } \mathbb{E}(|\psi'(\varepsilon_1)|) < \infty \text{ and such that the function } z \mapsto \mathbb{E}(|\psi'(\varepsilon_1 - z) - \psi'(\varepsilon_1)|) \text{ is}$$
continuous at zero.

(L4) \( \psi \) is a non-decreasing function, \( \mathbb{E}(\psi'(\varepsilon_1)) > 0 \) and \( \mathbb{E}(\psi'(\varepsilon_1)^2) < \infty \).

(\( \pi_2 \)) \( N^{-(1-mD)/2}||q_j|| \to \infty \), as \( N \to \infty \), for \( D < 1/m \), where \( m \) is the Hermite rank of the function \( \psi \circ G \), namely the smallest integer \( m \geq 1 \) such that \( \mathbb{E}[\eta\psi(G(\eta))] \neq 0 \), where \( H_m \) denotes the \( m \)th Hermite polynomial of order \( m \) and \( \eta \) is a standard Gaussian random variable.

Then, \( \hat{\beta}_N \) defined as solution of (11) satisfies

\[
D_N^{-1}(Q^TQ)(\hat{\beta}_N-\beta) = \{m! \mathbb{E}(\psi'(\varepsilon_1))\}^{-1} D_N^{-1} \sum_{i=1}^N q_{Ni} H_m(\eta_i) J_m + o_p \left( N^{(1-mD)/2} \right), \quad \text{as } N \to \infty .
\]

We shall also use the following theorem which is a reformulation of Theorems 1, 4 and 5 of Hurvich and Beltrán (1993).

**Theorem 6.2** Let \( (X_t) \) be mean-zero Gaussian process having a spectral density given by (7) and let

\[
A_j = \frac{1}{\sqrt{2\pi N}} \sum_{k=0}^{N-1} X_k \cos(k\lambda_j) , \quad B_j = \frac{1}{\sqrt{2\pi N}} \sum_{k=0}^{N-1} X_k \sin(k\lambda_j) ,
\]

where \( \lambda_j = 2\pi j/N \). Then, \( \mathbb{E}(A_j) = \mathbb{E}(B_j) = 0 \),

\[
\lim_{N \to \infty} \frac{1}{f(\lambda_j)} \mathbb{E}(A_j^2) = \mathcal{L}_1 , \quad \lim_{N \to \infty} \frac{1}{f(\lambda_j)} \mathbb{E}(B_j^2) = \mathcal{L}_2 , \quad \lim_{N \to \infty} \frac{1}{f(\lambda_j)^{1/2} f(\lambda_k)^{1/2}} \mathbb{E}(A_j B_k) = 0 ,
\]

where \( \mathcal{L}_1 \) and \( \mathcal{L}_2 \) are defined in (16) and (17), respectively.

**Proof of Theorem 2.7** We want to apply Theorem 6.1 with \( \beta = 0 \). For this, let us first check Condition (L1). Let \( \lambda_j = 2\pi j/N \) be a fixed Fourier frequency such that \( j \) belongs to \( \{1, 2, \cdots, [N/2]\} \). Let \( Q = C \), where \( C \) is defined by (5), then

\[
Q = \begin{pmatrix}
\cos(\lambda_j) & \sin(\lambda_j) \\
\cos(2\lambda_j) & \sin(2\lambda_j) \\
\vdots & \vdots \\
\cos(N\lambda_j) & \sin(N\lambda_j)
\end{pmatrix}
\]

and

\[
Q^TQ = \begin{pmatrix}
\sum_{k=1}^N \cos(k\lambda_j)^2 & \sum_{k=1}^N \cos(k\lambda_j) \sin(k\lambda_j) \\
\sum_{k=1}^N \cos(k\lambda_j) \sin(k\lambda_j) & \sum_{k=1}^N \sin(k\lambda_j)^2
\end{pmatrix} = \frac{N}{2} \text{Id}_2 ,
\]

which guarantees that (L1) holds. \( \text{Id}_2 \) denotes the \( 2 \times 2 \) identity matrix.

Let us now check Condition (\( \pi_2 \)). Observe that \( D_N = \sqrt{N/2} \text{Id}_2 \) and hence \( D_N^{-1} = \sqrt{2/N} \text{Id}_2 \). Thus

\[
d_i = (D_N)^{-1} q_{Ni} = \sqrt{2/N} (\cos(i\lambda_j) \sin(i\lambda_j))^T .
\]

Hence, \( N||d_i||^2 = 2 \), for all \( i \), which gives (\( \pi_1 \)).
Let us now check Condition (π2). Since \( q_{(j)} \) denote the \( j \)th column of \( Q \), \( \|q_{(1)}\| = \|q_{(2)}\| = \sqrt{N/2} \) and thus \( N^{-(1-mD)/2}\|q_{(j)}\| = N^{mD/2} \rightarrow \infty \), for all \( j \) and positive \( D \). Hence, (π2) is satisfied.

Observe that (A3) and (A4) in Theorem 6.1 correspond to (L2) and (L4) in Theorem 6.1. Moreover, (A5) implies that \( \mathbb{E}[\psi(\varepsilon_1)] = 0 \) and by (A2), Condition (1.3) of Theorem 6.1 is satisfied. Hence, Theorem 6.1 can be applied.

Let us now prove that the Hermite rank \( m \) of \( \psi_1 = \psi \circ G \) is equal to 1. Let \( \eta \) be a standard Gaussian random variable with p.d.f \( \varphi \), then

\[
J_1 = \mathbb{E}[\psi_1(\eta)\eta] = 2 \int_0^\infty y\psi_1(y)\varphi(y)\,dy ,
\]  

(31)

since \( \psi_1 \) is skew symmetric by (A5) and the assumption on \( G \). Note that \( J_1 \neq 0 \). Indeed, if \( J_1 = 0 \), since the integrand in (31) is nonnegative by the skew symmetric property of \( \psi_1 \), it would imply that \( \psi_1 = 0 \) which is in contradiction with (A2). Thus Theorem 6.1 applies with \( m = 1 \). To evaluate (28) recall that here: \( m = 1, H_1(\eta_k) = \eta_k \). Hence, by (30),

\[
\{m! \mathbb{E}(\psi'(\varepsilon_1))^\ast \}^{-1}(D_N)^{-1} \sum_{i=1}^N q_{Ni}H_m(\eta_i)J_m = \frac{J_1}{\mathbb{E}(\psi'(\varepsilon_1))} \sqrt{\frac{2}{N}} \sum_{i=1}^N \left( \frac{\cos(i\lambda_j)}{\sin(i\lambda_j)} \right) \eta_i .
\]

The asymptotic expansion (12) follows by using that \( (D_N)^{-1}Q^TQ = \sqrt{N/2} \mathbb{I}_d \).

Let us now prove (14). By (12), we have

\[
N^{D/2} \hat{\beta}_N(\lambda_j) = \frac{J_1}{\mathbb{E}(\psi'(\varepsilon_1))} \frac{2}{N^{1-D/2}} \sum_{i=1}^N \left( \frac{\cos(i\lambda_j)}{\sin(i\lambda_j)} \right) \eta_i + o_p(1) .
\]

The covariance of \( \frac{2}{N^{1-D/2}} \sum_{i=1}^N \left( \frac{\cos(i\lambda_j)}{\sin(i\lambda_j)} \right) \eta_i \) which is equal to

\[
\frac{4}{N^{2-D}} \sum_{1 \leq k, \ell \leq N} \begin{pmatrix} \cos(k\lambda_j) \cos(\ell\lambda_j) & \cos(k\lambda_j) \sin(\ell\lambda_j) \\ \cos(\ell\lambda_j) \sin(k\lambda_j) & \sin(k\lambda_j) \sin(\ell\lambda_j) \end{pmatrix} \rho(k - \ell) ,
\]  

(32)

converges to \( \tilde{\Gamma} \) defined in (14). Since the \( \eta_i \) are Gaussian, by Slutsky’s Lemma one gets (14).

To obtain (15), we need to compute the limit of (32) as \( N \) tends to infinity. Let \( U_{k,\ell}(\lambda_j) = \rho(k - \ell) \cos(k\lambda_j) \cos(\ell\lambda_j) \). Then,

\[
\frac{1}{f(\lambda_j)} \mathbb{E}(A_j^2) = \frac{1}{f(\lambda_j)} \frac{1}{2\pi N} \sum_{0 \leq k, \ell \leq N - 1} U_{k,\ell}(\lambda_j)
\]

\[
= \frac{1}{f(\lambda_j)} \frac{1}{2\pi N} \left[ \sum_{1 \leq k, \ell \leq N} U_{k,\ell}(\lambda_j) + 2 \sum_{1 \leq k, \ell \leq N - 1} U_{k,0}(\lambda_j) \right] .
\]

Since \( U_{k,0}(\lambda_j) \) is bounded by 1 in absolute value, the last sum in the last displayed expression above, normalized by \( 2\pi f(\lambda_j)N \), tends to 0 as \( N \) tends to infinity by (34).
Hence, by the first limit of (29) and (34),
\[
\lim_{N \to \infty} \frac{N^{-2d}}{(2\pi j)^{-2d} f^\star(0)} \frac{1}{2\pi N} \sum_{1 \leq k, \ell \leq N} \rho(k - \ell) \cos(k\lambda_j) \cos(\ell\lambda_j) = \mathcal{L}_1.
\]
Hence
\[
\lim_{N \to \infty} \frac{1}{N^{1+2d}} \sum_{1 \leq k, \ell \leq N} \rho(k - \ell) \cos(k\lambda_j) \cos(\ell\lambda_j) = 2\pi (2\pi j)^{-2d} f^\star(0) \mathcal{L}_1.
\]
By (10), we get
\[
\lim_{N \to \infty} \frac{4}{N^{2-D}} \sum_{1 \leq k, \ell \leq N} \rho(k - \ell) \cos(k\lambda_j) \cos(\ell\lambda_j) = 8\pi (2\pi j)^{-2d} f^\star(0) \mathcal{L}_1.
\]
Similarly, we obtain
\[
\lim_{N \to \infty} \frac{4}{N^{2-D}} \sum_{1 \leq k, \ell \leq N} \sin(k\lambda_j) \sin(\ell\lambda_j) \rho(k - \ell) = 8\pi (2\pi j)^{-2d} f^\star(0) \mathcal{L}_2.
\]
Moreover, by the third limit of (29), we get
\[
\lim_{N \to \infty} \frac{4}{N^{2-D}} \sum_{1 \leq k, \ell \leq N} \cos(k\lambda_j) \sin(\ell\lambda_j) \rho(k - \ell) = 0,
\]
which gives (15).

6.2. Proof of Corollaries

Proof of Corollary 2.5. Let \( \varphi \) be the p.d.f of a standard Gaussian random variable. By using an integration by parts and \( x \varphi(x) = -\varphi'(x) \), we get following (Giraitis et al., 2012, p. 24) that
\[
J_1 = \mathbb{E}[\eta \psi(\eta)] = \int_{\mathbb{R}} \psi(x) x \varphi(x) \, dx = \int_{\mathbb{R}} \psi'(x) \varphi(x) \, dx = \mathbb{E}[\psi'(\varepsilon_1)]. \tag{33}
\]
In view of Remark 2.2 and (13), both \( \hat{\beta}_N(\lambda_j) \) and \( \tilde{\beta}_{N,LS}(\lambda_j) \) have the same limit \( \mathcal{N}(0, \hat{\Gamma}) \).

Proof of Corollary 2.6. Since (18) yields
\[
N^{D-1} I_{N,\psi}(\lambda_j) = \frac{N^D}{8\pi} \| \hat{\beta}_N(\lambda_j) \|^2,
\]
Corollary 2.6 follows from the convergence in distribution given in (13) of Theorem 2.1.
Proof of Theorem 2.10

Since, by (9), as \( N \to \infty \),
\[
f(\lambda_j) \sim |\lambda_j|^{-2d} f^*(\lambda_j) \sim \left( \frac{2\pi j}{N} \right)^{-2d} f^*(0) = \left( \frac{2\pi j}{N} \right)^{D-1} f^*(0),
\]
(34)
one has
\[
\frac{I_{N,\psi}(\lambda_j)}{f(\lambda_j)} = \frac{N^{D-1} I_{N,\psi}(\lambda_j)}{N^{D-1} f(\lambda_j)} \rightarrow \frac{Z_i^2 + Z_2^2}{(2\pi j)^{-2d} f^*(0)} = \frac{J_1^2}{(\mathbb{E}[\psi'(\varepsilon_1)])^2} \left( L_1 W_1^2 + L_2 W_2^2 \right),
\]
(35)
by (19), (10), (20) and (15). By the continuous mapping theorem, the convergence in distribution of (35) also holds for the logarithm of these random variables, hence (22). \( \square \)

6.3. Proof of Theorem 2.10

It is enough to show that the Huber function defined in (23) satisfies Assumptions (A2) to (A5) since Assumption (A1) does not involve the function \( \psi \). We prove (A2), (A5), (A4), and (A3) successively.

By definition of \( \psi \), (A5) is satisfied. Let us now check (A2). Note that
\[
\mathbb{E}[\psi^2(\varepsilon_1)^2] = \mathbb{E}[\varepsilon_1^2 1_{|\varepsilon_1| \leq k}] + k^2 \mathbb{P}(|\varepsilon_1| > k) \leq k^2 \mathbb{P}(|\varepsilon_1| \leq k) + k^2 \mathbb{P}(|\varepsilon_1| > k) = k^2.
\]
Hence \( \mathbb{E}[\psi^2(\varepsilon_1)^2] < \infty \). By definition \( \mathbb{E}[\psi(\varepsilon_1)^2] \geq 0 \). If \( \mathbb{E}[\psi(\varepsilon_1)^2] = 0 \) then \( \mathbb{E}[\varepsilon_1^2 1_{|\varepsilon_1| \leq k}] = 0 \) and \( \mathbb{P}(|\varepsilon_1| > k) = 0 \). Thus, \( \mathbb{E}[\varepsilon_1^2] = 0 \). Since \( G(-x) = -G(x) \), we also have \( \mathbb{E}[\varepsilon_1] = 0 \) thus \( \varepsilon_1 = 0 \) almost surely, which is impossible. Since, by definition, \( \mathbb{E}[\psi(\varepsilon_1)^2] \geq 0 \), we thus get (A2).

Observe that \( \psi'(x) = 1_{|\varepsilon_1| \leq k} \). Since \( G(-x) = -G(x) \) and by the assumptions of (ii)', \( F_\varepsilon \) is continuous, we get \( F_\varepsilon(-t) = 1 - F_\varepsilon(t) \) and thus \( \mathbb{E}[\psi'(\varepsilon_1)] = 2F_\varepsilon(k) - 1 \). Using that \( F_\varepsilon(0) = 1/2 \) and that \( F_\varepsilon \) is increasing, \( \mathbb{E}[\psi'(\varepsilon_1)] > 0 \). Moreover, \( \mathbb{E}[\psi^2(\varepsilon_1)^2] = \mathbb{E}[\psi'(\varepsilon_1)] < \infty \), which gives (A1).

Observe that \( \mathbb{E}([\psi'(\varepsilon_1)]) = \mathbb{E}[\psi'(\varepsilon_1)] < \infty \). Also \( \mathbb{E}([\psi'(\varepsilon_1 - z) - \psi'(\varepsilon_1)]) = \mathbb{E}([1_{|\varepsilon_1 - z| \leq k} - 1_{|\varepsilon_1| \leq k}]) \). Since we will let \( z \to 0 \), it is enough to consider only the two cases \( 0 \leq z \leq 2k \) and \(-2k \leq z \leq 0 \). If \( 0 \leq z \leq 2k \), \( \mathbb{E}([\psi'(\varepsilon_1 - z) - \psi'(\varepsilon_1)]) = \mathbb{E}([1_{|\varepsilon_1| \leq z \leq k}]) = F_\varepsilon(z + k) - F_\varepsilon(k) \to 0 \), as \( z \to 0 \) by continuity of \( F_\varepsilon \) given by (ii'). If \(-2k \leq z \leq 0 \), \( \mathbb{E}([\psi'(\varepsilon_1 - z) - \psi'(\varepsilon_1)]) = \mathbb{E}([1_{-2k \leq \varepsilon_1 \leq -z}]) = F_\varepsilon(-k) - F_\varepsilon(z - k) \to 0 \), as \( z \to 0 \) by continuity of \( F_\varepsilon \) given by (ii'). This yields (A4) and concludes the proof.

Acknowledgments

Some empirical results presented in this paper were part of the PhD thesis of F.A. Fajardo at UFMG (Brazil) under supervision of V.A. Reisen. Both acknowledge the partial financial support from CNPq, FAPES and CAPES, Brazil. Murad S. Taqqu was supported by the NSF grant DMS-1300009 at Boston University. Part of this paper was revised when Prof. Valdério A. Reisen was visiting Centrale Supélec and AgroParistech (12/2016 to 01/2017). This author is indebted to Centrale Supélec for its financial support. The authors are grateful to the Editor and referees for the time and efforts in providing very constructive and helpful comments that have substantially improved the quality of the paper.
References


Reisen, V. A. and F. A. Fajardo (2012). Robust estimation in time series with long and