Penalised quantile periodogram for spectral estimation

Aymen Meziani, Tarek Medkour, Karim Djouani

Version of Record: https://www.sciencedirect.com/science/article/pii/S0378375818302489
Manuscript_e670cf72c17c5e802d459eca47ba4881

Penalised quantile periodogram for spectral estimation

Aymen Meziani[a,b], Tarek Medkour[b], Karim Djouani[a,c]

[a] *Laboratoire Images, Signaux et Systèmes Intelligents (LiSSi), Université de PARIS-EST, Paris, France*
[b] *Department of Statistics and Probability, University of USTHB, Algiers, Algeria*
[c] *Tshwane University of Technology, FSATI, Pretoria, South Africa*

Abstract

The present paper introduces an extension of the quantile periodogram developed in [1] since the latter suffers from instabilities and spectral leakage. To overcome these drawbacks, a L_1 regularised version is proposed. The asymptotic properties of the new spectral estimator are discussed and simulations are performed in order to show its efficiency in the presence of hidden periodicities and under different types of noise.

Keywords: Periodogram, Quantile regression, LASSO penalisation, spectral density estimation, asymptotic distribution, Hidden periodicity

1. Introduction

Spectral analysis techniques aim to quantify the energy generated over the frequency components of a given time series. The periodogram has commonly been used as a non-parametric estimator of the spectrum. However, it may not exhibit robustness, especially in the presence of heavy-tailed data or outliers; see for example [2].

The ordinary periodogram uses least squares estimates for the projection of the time series onto the Hilbert space generated by the sinusoidal base. The least squares estimates focus only on the conditional mean. The quantile regression introduced in [3] attempts to evaluate the variation of the conditional quantiles with respect to the response variable. The quantile regression covers the whole distribution of the data and exhibits robustness against outliers. The quantile periodogram introduced in [1] summarises the impact of the cyclical behaviour on the distribution of the time series. The random coefficients autoregressive models, as described in [4], are an example for the application of the quantile periodogram.

The quantilogram introduced in [5], has been used to evaluate the directional predictability in the time domain. The quantile spectrum studies introduced in [4] and [6] use the level crossings for spectral estimation instead of the time series. Based on the spectral density, the authors in [7] propose a test for the sequential dependence of non-linear time series. The composite quantile periodogram is proposed in [8] in order to overcome the problem of the choice of the quantile of interest.

In order to improve the prediction accuracy and to reduce the variance of estimators, different types of penalisation methods have been introduced. The L_q penalisation family is among the most commonly used, with a wide interest in the L_1 penalty, such as the LASSO penalty (see, e.g., [9]) and the L_2 penalty used in ridge regression. Since the L_q penalisation family does not satisfy the Oracle properties, different types of penalisation functions have been proposed in [10] and [11].

The L_1 norm was used in the classical spectral estimation in [12] and [13], where the authors obtained robust estimators against leakage for noisy and uneven data.

The present paper discusses a regularised version of the quantile periodogram proposed in [1] to overcome its spectral leakage issue. Based on the L_1 norm, the penalised quantile periodogram could be used as a sparse representation of the quantile spectra. The proposed estimator exhibits a good frequency domain representation of serial dependence. In addition, it can reduce the spectral leakage produced by the quantile periodogram as it shrinks some elements towards zero. This is illustrated using the Fisher test for processes with hidden periodicities. It is also demonstrated that the penalised quantile periodogram is related to the level-crossing spectrum. Thus, it shares the invariance to non-linear distortions.

The paper is organised as follows. A brief introduction to classical spectral analysis and the quantile periodogram is provided in sections 2 and 3 respectively; putting emphasis on definitions required for further developments. Our contribution is introduced in section 4, where the penalised quantile periodogram is described. The asymptotic theory of the new estimator is presented in section 5. In section 6, comparative study on the performances of the different spectral estimators on simulated as well as real data is presented. The conclusion and future work are discussed in the last section.

2. Classical Spectral Analysis

Any real-valued time series $\{Y_1, Y_2, ..., Y_n\}$ can be written using a Fourier representation as follows (see, e.g., [14])

$$Y_t = \lambda + \sum_{j=1}^{q} \beta_{1j} \cos(\omega_j t) + \beta_{2j} \sin(\omega_j t), \tag{1}$$

with $\omega_j = 2\pi j/n$, $j = 1, .., q$, the Fourier frequencies and $q = [n/2]$, while $(\lambda, \beta_{1j}, \beta_{2j}) \in \mathbb{R}^3$ are the coefficients to be estimated. Commonly used as a non-parametric estimator of the spectral density, the periodogram is expressed as follows

$$I(\omega_j) = \frac{1}{n} \left| \sum_{t=1}^{n} Y_t \exp(-it\omega_j) \right|^2 .$$

The last expression can be written in an equivalent manner using (1) as follows

$$I(\omega_j) = \frac{1}{4} n(|\beta_{1j}|^2 + |\beta_{2j}|^2) = \frac{1}{4} n \left\|\beta_j\right\|_2^2 ,$$

with $\beta_j = [\beta_{1j}, \beta_{2j}]^T$. The coefficients are estimated using a linear least squares estimator of the form

$$\widehat{\beta}(\omega_j) = \arg\min_{\beta \in \mathbb{R}^2} \left\| Y_t - x^T(\omega_j)\beta \right\|_2^2 ,$$

where $x(\omega_j) = [\cos(\omega_j), \sin(\omega_j)]^T$ are the harmonic regressors. The periodogram is not a consistent estimator of the spectrum. Furthermore, it requires the existence of the second moment to have a well defined asymptotic distribution, as detailed in [2].

3. Quantile Periodogram

Least squares techniques result in the estimation of the conditional mean of the response variable given the predictors. Quantile regression techniques, however, focus on the estimation of quantiles of the response variable. The quantile periodogram introduced in [1] demonstrated an effective performance in detecting hidden periodicities in time series providing a complete view of the effect of the periodicities in each part of the distribution of the data. Furthermore, it is robust against heavy-tailed noise and non-linear

distortions. Using the quantile regression to estimate the coefficients in (1) gives

$$\widehat{\lambda} = \arg\min_{\lambda} \sum_{t=1}^{n} \rho_\tau(Y_t - \lambda),$$

and

$$\widehat{\beta}_\tau(\omega_j) = \arg\min_{\beta \in \mathbb{R}^2} \sum_{t=1}^{n} \rho_\tau(Y_t - \lambda - x_t^T(\omega_j)\beta), \tag{2}$$

with $\widehat{\beta}_\tau(\omega_j) = [\widehat{\beta}_{\tau,1}(\omega_j), \widehat{\beta}_{\tau,2}(\omega_j)]$ and $x_t(\omega_j) = [\cos(\omega_j), \sin(\omega_j)]^T$, while λ is a suitable constant, typically the τ-quantile of $\{Y_t\}$.

$\rho_\tau(u)$ is the loss function defined by

$$\rho_\tau(u) = u\{\tau - I(u < 0)\} = \begin{cases} -1(1-\tau)u, & \text{if } u < 0, \\ \tau u, & \text{if } u \geq 0, \end{cases}$$

where $I(.)$ is the indicator function and τ is the quantile level. Therefore, given the following expression

$$\widetilde{Y}_t = \widehat{\lambda} + \sum_{j=1}^{q} \widehat{\beta}_{\tau,1}(\omega_j)\cos(\omega_j t) + \widehat{\beta}_{\tau,2}(\omega_j)\sin(\omega_j t), \tag{3}$$

the quantile periodogram (QP) of Y_t is defined as follows

$$Q_\tau(\omega_j) = \frac{1}{4} n \left\| \widehat{\beta}_\tau(\omega_j) \right\|_2^2. \tag{4}$$

Theorem 2 in [1] pointed out that when estimating the quantile periodogram at a given frequency ω_0, spectral leakage may occur at other frequencies (e.g., the integral multiples of ω_0), especially, when the periodicity is strong. This leakage is even stronger in the presence of multiple periodicities, affecting significantly the estimation at some frequencies. In order to overcome such a drawback, a regularised version of the quantile periodogram is proposed.

4. Penalised Quantile Periodogram

The penalisation methods are a class of algorithms that transform a constraint optimisation problem into an unconstrained one by adding a term called the penalty function.

In order to overcome the problem of leakage exhibited by the original quantile periodogram estimator, a penalised quantile periodogram is proposed. It is based on the L_1 penalty, that shrinks some of the predictors towards zero.

For regression problems, the notion of the Oracle properties is related to the consistency of both parameter estimation and variable selection (unbiasedness, sparsity and continuity). The non satisfaction of the Oracle properties may result in inconsistency in parameter estimation, variable selection or both, as discussed in [15]. The authors in [10] show that none of the L_q penalisation family satisfy the Oracle properties. In [11], a generalised definition of the LASSO penalty is introduced using a weight function. The authors proposed to use the reciprocal of the ordinary least squares estimates raised to some power. For quantile regression, the authors in [15] proposed the utilisation of the quantile regression estimates as weights.

The penalised quantile periodogram is defined by adding a penalisation term to expression (2). The LASSO quantile periodogram based on the typical L_1 norm is defined as follows

$$Q^{(L)}_{\tau,\alpha_n}(\omega_j) = \frac{n}{4}\left\|\widehat{\beta}^{(L)}_{\tau,\alpha_n}(\omega_j)\right\|_2^2, \tag{5}$$

where

$$\widehat{\beta}^{(L)}_{\tau,\alpha_n}(\omega_j) = \arg\min_{\beta\in\mathbb{R}^2}\sum_{t=1}^{n}\rho_\tau(Y_t - \lambda - x_t^T(\omega_j)\beta) + \alpha_n\left\|\beta\right\|_1. \tag{6}$$

$\|.\|_1$ represents the L_1 norm, $\alpha_n \geq 0$, is the regularisation parameter and $\|\beta\|_1$ denotes the penalty function of β.

In order to overcome the inconsistency problem exhibited by the LASSO quantile periodogram, the *adaptive LASSO quantile periodogram* (ALQP) is introduced as follows

$$Q^{(AL)}_{\tau,\alpha_n}(\omega_j) = \frac{n}{4}\left\|\widehat{\beta}^{(AL)}_{\tau,\alpha_n}(\omega_j)\right\|_2^2, \tag{7}$$

with

$$\widehat{\beta}^{(AL)}_{\tau,\alpha_n}(\omega_j) = \arg\min_{\beta\in\mathbb{R}^2}\sum_{t=1}^{n}\rho_\tau(Y_t - \lambda - x_t^T(\omega_j)\beta) + \alpha_n\sum_{i=1}^{2}\vartheta_i|\beta_i|, \tag{8}$$

where $\vartheta_i = 1/|\widehat{\beta}^{(i)}_\tau|^d$ for some $d > 0$ and $\widehat{\beta}_\tau$ is as defined in (2).

Let $A_j^{(n)} = \{i : \widehat{\beta}_{\tau,\alpha_n}(\omega_j) \neq 0\}$ be the subset of non zero parameters at each frequency. The variable selection procedure is consistent if and only if $\lim_n P(A_j^{(n)} = A_j) = 1$, where A_j is the correct subset model. The Oracle properties of (8) are demonstrated in the appendix. The asymptotic properties of the ALQP estimator are detailed below.

5. Asymptotic Properties

In this section, several theoretical results concerning the asymptotic distribution of the ALQP defined in equation (7) are discussed.

Let $\{Y_t\}$ be a time series satisfying the assumptions in Theorem 1 (see below). Let $S(\omega)$ be the level-crossing spectrum defined by

$$S(\omega) = \sum_{h=-\infty}^{\infty} \left\{1 - \frac{1}{2\tau(1-\tau)}\gamma_h\right\} \cos(\omega h), \tag{9}$$

where γ_h is the lag-h level crossing rate of $\{Y_t\}$ defined by

$$\gamma_{t-s} = P\{(Y_t - \lambda)(Y_s - \lambda) < 0\}.$$

The level-crossing spectrum can be regarded as an alternative spectral representation of the serial dependence of a time series. It has been proven in [1] that there exists a non-trivial transformation linking the level crossing rate to the autocorrelation function under the Gaussian white noise assumption. Therefore, it is possible to establish a direct relationship between the level crossing spectrum and the ordinary power spectrum.

It is well known that the ordinary periodogram is related to the autocorrelation spectrum. The quantile periodogram; similarly, is found to have a relationship with the so-called level-crossing spectrum. In the next theorem, we show that the ALQP preserves the same relationship using the adaptive LASSO penalisation.

Theorem 1. *Let $F_t(u)$ and $F_{t,s}(u,v)$ denote the univariate and bivariate marginal distribution functions of $\{Y_t\}$. Assume the following conditions are satisfied for a fixed τ and λ*

(C1) $f_t(u) := F_t'(u)$ exists for all u and $F_t(u+\lambda) - F_t(\lambda) = f_t(\lambda)u + O(u^{d+1})$ uniformly for $|u| \leq u_0$.

(C2) $F_t(\lambda) = \tau$ and $f_t(\lambda) = \kappa > 0$ for all t.

(C3) $\{Y_t\}$ *is stationary in* λ*-level crossings, that is* $P\{(Y_t - \lambda)(Y_s - \lambda) < 0\} = \gamma_{t-s}$ *for all* t *and* s.

(C4) $\{Y_t\}$ *is an* m*-dependent process or a linear process of the form* $\lambda + \sum_{l=-\infty}^{+\infty} \phi_l e_l$*, where* $\{e_t\}$ *is an iid random sequence with* $E(|e_t|) < \infty$ *and* $\{\phi_l\}$ *is an absolutely summable deterministic sequence such that*

$$\sum_{|l|>n^r} \phi_l = o(n^{-1}) \quad as \quad n \to \infty,$$

for some constant $r \in [0, \frac{1}{4}]$.

Given $Q_{\tau,\alpha_n}^{(AL)}$ *defined by (7), assuming that* $S(\omega)$ *is finite and bounded away from zero and also assuming that* A_j *is the correct subset model for each frequency, let* p_j *be the length of* A_j*. Then, because* $\alpha_n/\sqrt{n} \to 0$ *and* $\alpha_n n^{(d-1)/2} \to \infty$*, we have*

$$\{Q_{\tau,\alpha}^{(AL)}(\omega_j)\} \overset{A}{\sim} \left\{ \frac{1}{2}\eta^2 S(\omega_j)(\xi_{p_j}) \right\},$$

with $\eta^2 = \tau(1-\tau)/\kappa^2$ *and* $\xi_{p_j} \sim \chi^2_{p_j}$*,* $p_j \in \{1, 2\}$.

See the appendix for the detailed proof of this theorem.

Theorem 1 states that the ALQP has a similar asymptotic behaviour to the QP, in the sense that they are both proportional to the quantile spectra ($\eta^2 S(\omega_j)$ as defined in [1]). The difference arises with respect to the limiting distribution, which is a chi-square with two degrees of freedom for all frequencies in the case of the QP, whereas, in the case of the ALQP, it could be a chi-square with one degree of freedom for some frequencies.

It is to be noted that at some frequencies, all the estimated parameters could be shrunk towards zero. Therefore, in our context, the sparsity pertains to the selection of the harmonic regression parameters. Explicitly, this implies that there could be some frequencies for which the ALQP is exactly zero.

Due to the association of the quantile estimators with the level-crossing spectrum, they are invariant to non-linear distortions. This means that any non-linear memoryless transformation of the data affects only the scaling constant in the asymptotic distribution. The ordinary periodogram does not have this advantage, as non-linearity will deform the entire power spectrum in a non-trivial way (see, e.g., [1, 16]). Furthermore, the level-crossings are known to have the ability to provide richer information about the serial dependence structure of time series.

6. Numerical results

In this section, we provide some numerical results to demonstrate the efficiency of the ALQP and to confirm the theoretical conclusions discussed in the previous section. We demonstrate that the penalised quantile periodogram is a powerful tool for detecting hidden periodicities in different parts of the distribution of the data, even with the existence of multiple periodicities and under non-standard assumptions about the noise.

The choice of the tuning parameter α is very crucial because it has a strong relationship with the sparsity of the model. There are many tools used in the literature for choosing an optimal tuning parameter, such as generalised cross-validation, the Akaike information criterion and the Bayesian information criterion. In [17], the authors show that using the adaptive LASSO penalty and the BIC tuning parameter selector can be as efficient as the Oracle.

In this paper, the parameter α is chosen using the minimisation of the BIC criterion for each frequency in a range of $[10^{-1}, 10^{2}]$

$$BIC = \log(\frac{1}{n}\sum_{t=1}^{n}\rho_\tau(Y_t - \beta^T x_t)) + \frac{\log(n)}{n} edf,$$

where *edf* is the number of non zero coefficients in $\hat{\beta}^{(AL)}_{\tau,\alpha_n}$. The procedure to estimate α is performed for each frequency separately. We can also choose to use a unique value by selecting the mean over all frequencies, for example. Cross-validation techniques could also be used in this context. Nevertheless, an automatic procedure for the detection of the minimum regularisation parameter is more indicated. The sensitivity of the ALQP to the choice of the regularisation parameter is also discussed.

For simulation purposes, the following model is used

$$\begin{aligned} Y_t &= a_t X_t, \\ a_t &= b_0 + b_1\cos(\omega_0 t) + b_2\sin(\omega_1 t), \end{aligned} \tag{10}$$

with ($b_0 = 1$, $b_1 = 0.9$, $b_2 = 1$) and $\{X_t\}$ is an AR(2) process, satisfying

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \xi_t \quad (t = 1, \ldots, n), \tag{11}$$

with $\phi_1 = 2r\cos(\omega_c)$ and $\phi_2 = -r^2$ ($r = 0.6$, and $\omega_c = 2\pi * 0.22$), while $\{\xi_t\}$ is the noise.

We use $\omega_0 = 2\pi \times 0.1$ and $\omega_1 = 2\pi \times 0.12$. The parameter d mentioned in equation (8) is chosen to be equal to 2. The goal is to test the ability of the different estimators to detect multiple neighbouring periodicities. Different simulations using different types of noises are used to compare the robustness of the estimators against non-standard settings.

Like the ordinary periodogram and the quantile periodogram, the ALQP could also serve as a representation of serial dependence in the frequency domain for time series without hidden periodicities, as shown in Figure 1.

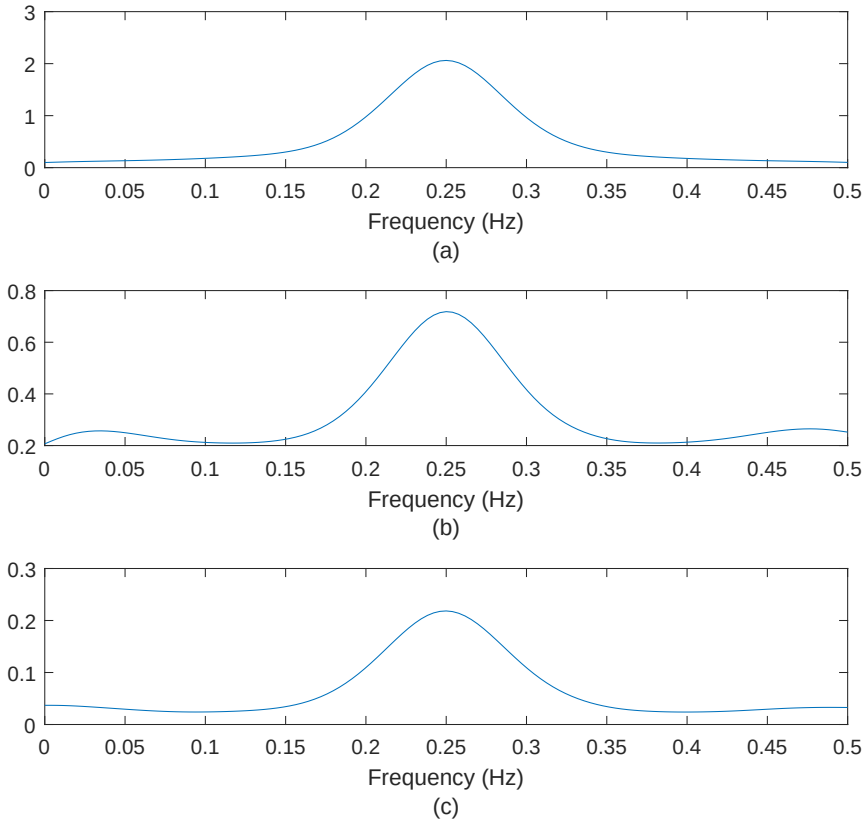

Figure 1: Ensemble mean of 1000 smoothed periodograms of the AR(2) time series defined in equation (11) with standard white noise. (a) shows the ordinary periodogram, (b) the quantile periodogram and (c) the ALQP. The sample size used is $n = 200$. The quantile value used is $\tau = 0.8$.

For all estimators, the smoothing was performed using a smoothing spline function with the same parameter. The ALQP behaves similarly to the quantile and ordinary periodograms exhibiting a "bell-shaped" pattern, as expected for a power spectrum of an AR(2) process.

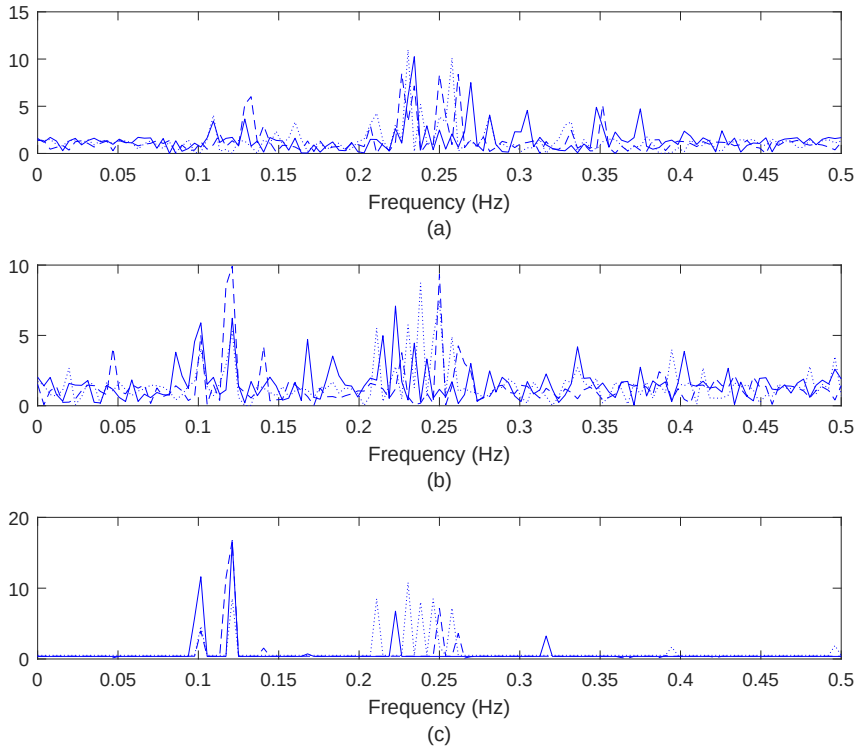

Figure 2: Spectral density estimators for three types of noise for a single realisation of the model defined by (10). (a) shows the ordinary periodogram, (b) the quantile periodogram and (c) the ALQP. The dotted line (.) represents the standard Gaussian noise, the dashed line (-) the large variance Gaussian noise and the solid line the student-t noise. The sample size used is $n = 200$. The quantile value used is $\tau = 0.8$.

In a first experiment, the estimators considered were obtained for a single realisation for the time series model (10) with different added noise types, as depicted in Figure 2. A quantile value $\tau = 0.8$ was used for the quantile estimators. It is clear that only our newly introduced estimator detects the periodicities with different types of noises. For the quantile periodogram

estimator, the presence of multiple strong periodicities affects significantly the estimation of the spectrum at some frequencies.

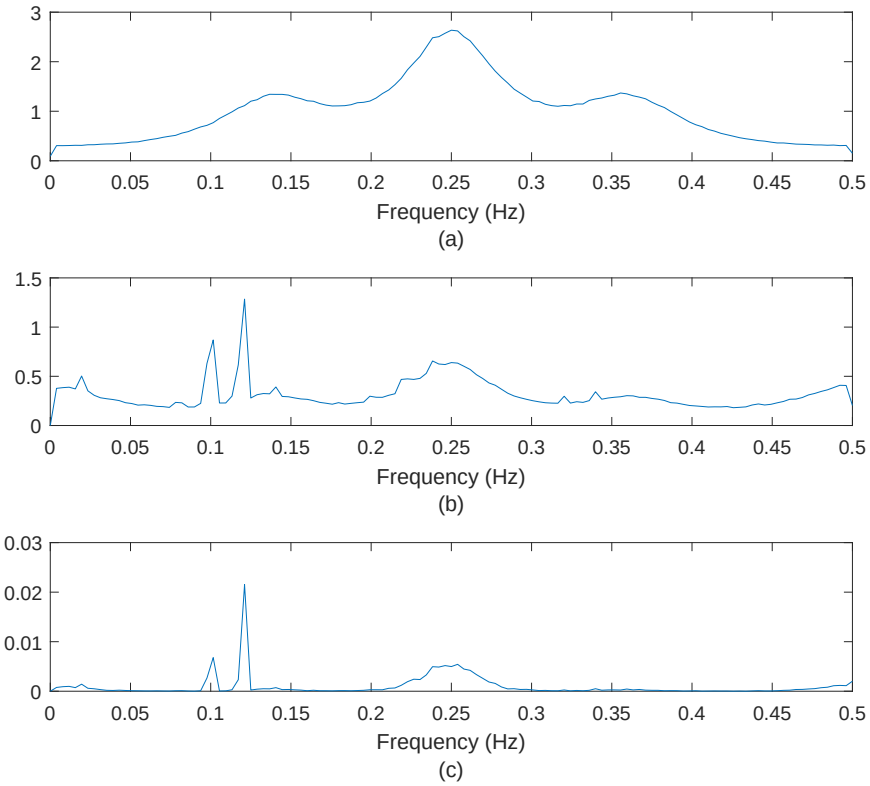

Figure 3: Mean of 5000 realisations of estimated spectral densities using student-t noise. (a) represents the ordinary periodogram, (b) the quantile periodogram and (c) the ALQP. The sample size used is $n = 200$. The quantile value used is $\tau = 0.8$.

Figure 3 depicts a mean of 5000 realisations of the spectral estimators for the time series model defined in (10) with student-t noise and using $\tau = 0.8$ for the quantile estimators. Both the QP and the ALQP detect the correct periodicities. The quantile periodogram, however, suffers from random spikes which occur next to the true periodicities and elsewhere, as suggested by Theorem 2 in [1].

Using the ordinary periodogram, a hypothesis test for the detection of

hidden periodicities could be proposed. The Fisher g test defined by

$$g = \frac{max_{1\leq k\leq m}(I_n(\omega_k))}{\sum_{k=1}^{m} I_n(\omega_k)}, \quad m = [n/2], \tag{12}$$

is used. The aim is to test the null hypothesis that the time series is a Gaussian white noise against the alternative hypothesis that the time series contains a periodic component. If the periodogram contains a value substantially larger than the average value (i.e., if g is sufficiently large), the null hypothesis is rejected. This test was applied in [1] by replacing the ordinary periodogram in (12) with the quantile periodogram. The same concept could be applied here as well for the ALQP under the assumption that the ALQP of a white noise has a scaled χ^2_2 distribution independent of the frequency ω_j (see [18]). This is clearly evident in Figure 4, where the distributions of the QP and the ALQP estimators for white noise data are compared. Figure 4 represents the cumulative distribution, at a chosen frequency, of the ALQP and the QP for 1000 simulated standard white noise samples.

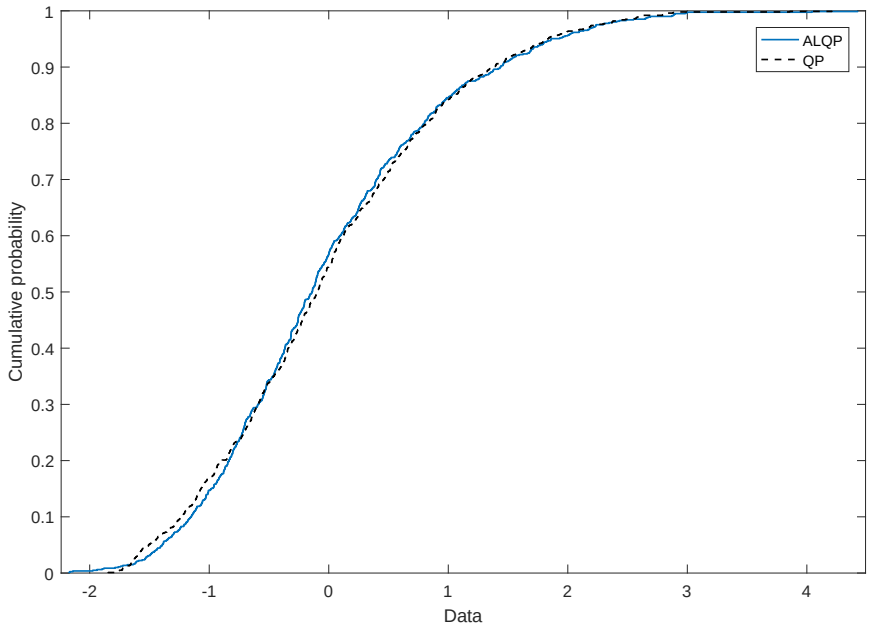

Figure 4: The cumulative distribution, at a fixed frequency, of the ALQP and the QP estimated from 1000 simulated standard white noise samples.

The null hypothesis is rejected if the P-value is less than $\widetilde{\alpha}$ at the significance level $\widetilde{\alpha}$.

Significance level	Periodogram	Qp			ALQP		
		$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
0.01	0	0.2380	0.5100	0.2120	0.8410	0.9730	0.8480
0.05	0.01	0.3950	0.6870	0.4150	0.8700	0.9770	0.8680
0.1	0.02	0.4950	0.7680	0.5080	0.8810	0.9790	0.8680

Table 1: Power of Fisher's test for 1000 Monte Carlo simulations of the time series defined in the first example using one periodicity ($\omega_0 = 0.1 \times 2\pi$). The sample size used is $n = 200$.

Table 1 indicates the power of the test calculated using the g statistic defined in (12). 1000 Monte Carlo simulations were performed for the three spectral estimators of the time series defined in (10) with a single periodicity at $\omega_0 = 0.1 \times 2\pi$ (i.e., $b_2 = 0$) and standard Gaussian noise. The results show that at level 0.01, the ALQP has a 97.3% chance of detecting the periodicity using $\tau = 0.8$, whereas the quantile periodogram has only a 51% chance at the same quantile level. We note that the ALQP is not as sensitive to the choice of the quantile value τ as the quantile periodogram. Nevertheless, a proper choice of τ is necessary to maximise the power of the test.

For the analysis of the sensitivity of the ALQP to the choice of the regularisation parameter, the same experiments were performed with different values of α. Table 2 indicates the power of the test for 1000 simulations of the ALQP with a unique α for all frequencies and $\tau = 0.8$. Note that the detection power is also α dependent. However, a large interval of values records similar performances. Therefore, to maximise the power detection, a proper range of α is required.

Significance level	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
0.01	0	0.0030	0.8990	0.9830
0.05	0.02	0.0150	0.9630	0.9870
0.1	0.02	0.0410	0.9780	0.9890

Table 2: Power of Fisher's test for the ALQP estimates with different regularisation values. 1000 simulations used with the quantile value $\tau = 0.8$. The sample size used is $n = 200$.

Fisher's test is only concerned with the largest value of the periodogram ordinates, although it can be replaced with the true periodicity. An extended

Significance level	Periodogram		QPeriodogram		ALQP	
	1st per	2nd per	1st per	2nd per	1st per	2nd per
0.01	0.0000	0.0000	0.0440	0.0500	0.6600	6290
0.05	0.0000	0.0010	0.1160	0.1120	0.6930	6480
0.1	0.0010	0.0010	0.1580	0.1630	0.7130	0.6970

Table 3: Power of Fisher's test for 1000 Monte Carlo simulations of the time series defined in the first example using two periodicities ($\omega_0 = 0.1 \times 2\pi$, $\omega_1 = 0.12 \times 2\pi$).

version of Fisher's test was therefore proposed in [19] for the detection of a second peak by omitting the term of the largest peak from the denominator of the g statistic and substituting $(n-1)$ for n in the distribution of the latter. Thus, if the largest peak turns out to be significant, we can test for a second peak using

$$g' = \frac{I_n(\omega_1)}{\{\sum_{k=1}^{m} I_n(\omega_k)\} - \max_{1\leq k\leq m}(I_n(\omega_k))}. \tag{13}$$

The same procedure can be repeated until the next largest ordinate is not significant in the sense of the considered test.

Table 3 depicts the results of 1000 Monte Carlo simulations of the test statistic defined in (13) for the three spectral estimators of the time series defined in (10) with two periodicities. The obtained results show that at level 0.1 the ALQP has a 71.3% chance of detecting the first periodicity and a 69.7% chance of detecting the second one, whereas, the quantile periodogram has only a 15.8% chance for the first one and a 16.3% chance for the detection of the second periodicity.

Real data

The yearly sunspot dataset is used to illustrate the performance of the different estimators in the case of a real example. This dataset is known to have an 11-year cycle. In [1], the quantile periodogram was shown to be able to detect a clear peak at 0.09=1/11, at lower quantiles. The ordinary periodogram had extra peaks on both sides of the real periodicity as a result of spectral leakage. In our case, we added a student-t noise to the dataset, as in Figure 5.

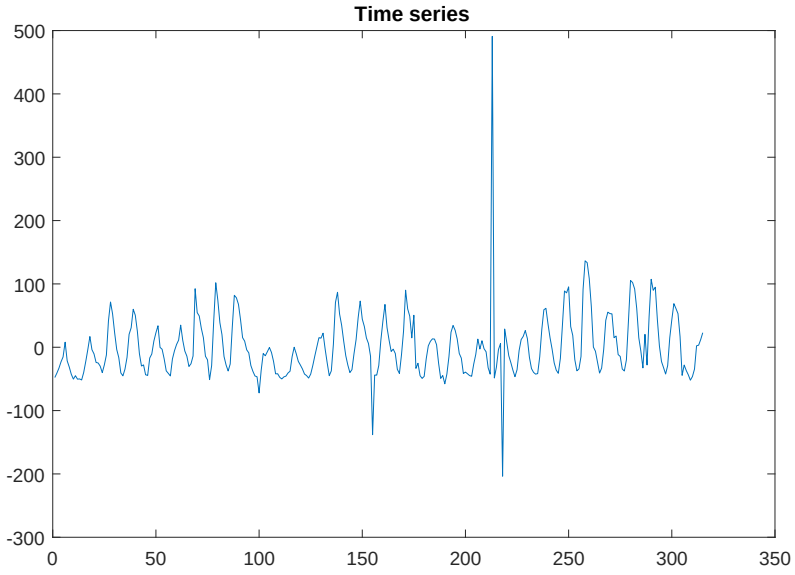

Figure 5: Yearly sunspot numbers time series with added student-t noise.

Figure 6 depicts a single realisation of the different spectral estimators. The quantile value used is 0.25 as it presented the best results for the quantile estimators. The regularisation parameter was estimated using the BIC criterion. We can observe the ability of the ALQP to detect the only and true periodicity.

7. Conclusion

A new frequency domain estimator is proposed by augmenting the quantile periodogram with the adaptive LASSO penalisation term. This choice has been motivated by the presence of spectral leakage in the estimation of the quantile periodogram. We have demonstrated that the proposed estimator is related to the level-crossing spectrum and, consequently, shares the property of invariance to non-linear distortions. In addition, it could serve as a viable serial dependence representation in the frequency domain. Our simulations demonstrate that the penalised quantile periodogram outperforms the other spectral estimators in terms of detecting hidden periodicities, and is robust against deviation from standard white noise. However, future research is required to develop methods to determine the appropriate values of

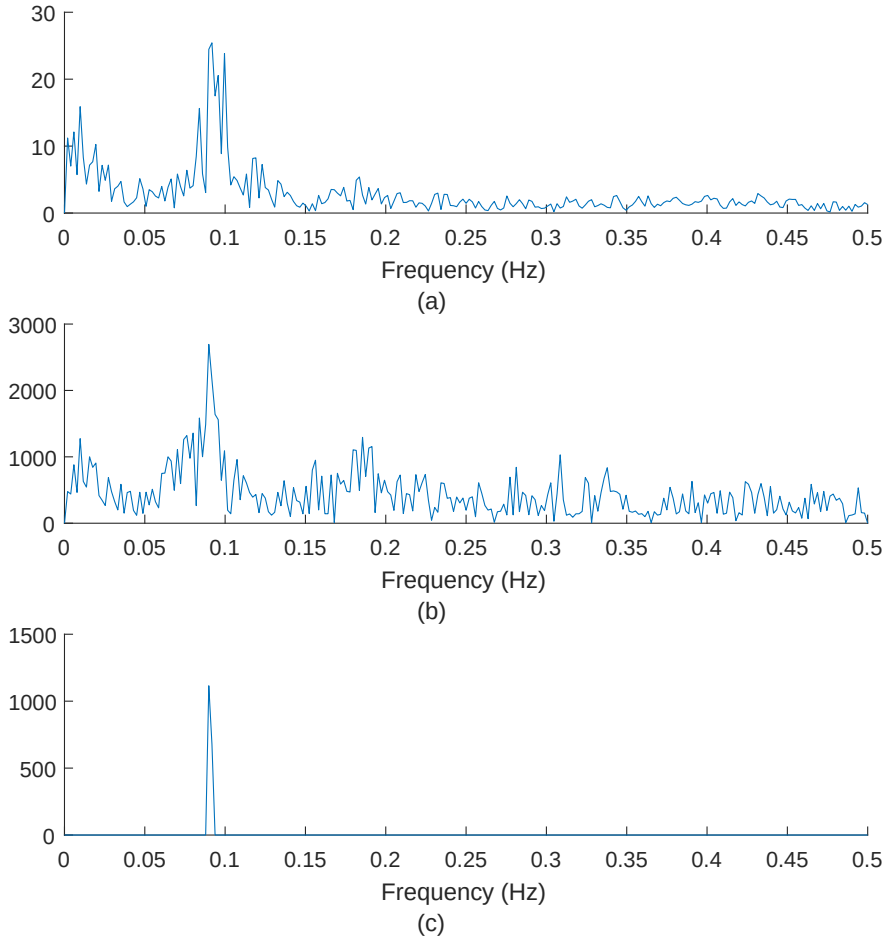

Figure 6: Spectral estimators for the yearly sunspot numbers with added student-t noise. (a) represents the ordinary periodogram, (b) the quantile periodogram and (c) the penalised quantile periodogram.

the quantile and the regularisation terms.

Acknowledgements

The work of the third author is based on the research supported in part by the National Research Foundation of South Africa (Grant Number: 90604). Opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are those of the author(s) alone, and the NRF accepts no liability whatsoever in this regard.

8. Appendix

8.1. Penalised Quantile Regression Theorem

For the proof of the main theorem, the asymptotic theory of the penalised quantile regression problem is discussed. With respect to the work that has been done in [1] and [18], a more general problem of penalised quantile regression is considered. For each $j = 1, .., q$, let $\{x_{jt}\}$ be a bounded sequence of deterministic vectors in $\mathbb{R}^p$ for some $p \in \mathbb{N}$.

Let Y_t take the form $Y_t = u_t + \varepsilon_t$ where $\{u_t\}$ is a deterministic sequence and $\{\varepsilon_t\}$ is an m-dependent random process having the marginal distribution functions $G_t(u)$, marginal densities $g_t(x) = G_t'(x)$, while $G_{t,s}(u, v)$ denotes the bivariate marginal distribution function. Let $r_{ts}(u, v) = G_{ts}(u, v) - G_t(u)G_s(v)$. For any constant vector $\beta_j \in \mathbb{R}^p$, let $w_{jt} = x_{jt}^T\beta_j - u_t$.

Assume that there are positive numbers d, c and N_0 and positive definite matrices C_0 and Λ_0. In addition, let $A_j = \{i : \beta_{j,i} \neq 0\}$, be the right subset model and assume without loss of generality that $A_j = \{1, 2, .., p_j\}$, $p_j \leq p$ such that

(i) $g_t(w_{jt}) = O(1)$ uniformly in $t \in \mathbb{N}$ for all j.

(ii) $G_t(u_t + w_{jt}) - G_t(w_{jt}) = g_t(w_{jt})u + O(u^{a+1})$ uniformly in $t \in \mathbb{N}$ for $|u| \leq c$ and all j, $a > 0$.

(iii) $C_{jkn} = n^{-1}\sum_{t=1}^n \sum_{s=1}^n r_{ts}(w_{jt}, w_{ks})x_{jt}x_{ks}^T \geq C_0$ for all j, k and $n \geq N_0$.

(iv) $\Lambda_{jn} = n^{-1}\sum_{t=1}^n g_t(w_{jt})x_{jt}x_{jt}^T \geq \Lambda_0$ for all j and $n \geq N_0$.

Under these conditions, the following results can be established. For instance, let $\widehat{\beta}_{j,\alpha_n}^{(L)}$ be defined by

$$\widehat{\beta}_{j,\alpha_n}^{(L)} = \arg\min_{\beta\in\mathbb{R}^p} \sum_{t=1}^n \rho_\tau(Y_t - \lambda - x_{jt}^T\beta) + \alpha_n \sum_{i=1}^p |\beta_i|.$$

Lemma 1. *If* $\alpha_n/\sqrt{n} \to 0$ *and* $\alpha_n \to \infty$, *then*

$$\frac{n}{\alpha_n} vec[\widehat{\beta}^{(L)}_{j,\alpha_n} - \beta_j]^q_{j=1} \underset{p}{\to} \arg\min(V_n(\delta)),$$

where

$$V_n(\delta) = -\delta^T \zeta_{jn} + \frac{1}{2}\delta^T \Lambda_{jn} \delta + \alpha_{\mathbf{n}} \sum_{i=1}^{p} \left[\delta_i sgn(\beta_i) I(\beta_i \neq 0) + |\delta_i| I(\beta_i = 0)\right].$$

ζ_{jn} *has a* $\mathcal{N}(h_{jn}, C_{jkn})$ *distribution.*

Remark 1. The LASSO penalisation has been shown to be an effective estimator only under some restrictive conditions. To overcome this problem, the adaptive LASSO penalty proposed in [11] is used.

Thus, the Oracle properties of the ALQP problem are established in the special case presented in this paper. Assume that $\widehat{\beta}^{(AL)}_{j,\alpha_n}$ defined in (8), can be written as $\widehat{\beta}^{(AL)}_{j,\alpha_n} = [\widehat{\beta}^{(AL)}_{j,\alpha_n 1}, \widehat{\beta}^{(AL)}_{j,\alpha_n 2}]$ where $\widehat{\beta}^{(AL)}_{j,\alpha_n 1} \in \mathbb{R}^{p_j}$ and $\widehat{\beta}^{(AL)}_{j,\alpha_n 2} \in \mathbb{R}^{p-p_j}$.

Theorem 2. . *Assuming that (i) to (iv) are satisfied and that* $\alpha_n/\sqrt{n} \to 0$ *and* $\alpha_n^{(d-1)/2} \to \infty$, *then we have*

$$\sqrt{n} vec[\widehat{\beta}^{(AL)}_{j,1} - \beta_{j,1}]^q_{j=1} \overset{A}{\sim} \mathcal{N}(\theta_{A_j}, \Sigma_{A_j}),$$

where $\Sigma_{A_j} = [\Lambda^{-1}_{j,A_j} C_{j,A_j} \Lambda^{-1}_{j,A_j}]^q_{j=1}$, $\theta_{A_j} = vec\{\Lambda^{-1}_{j,A_j} h_{j,A_j}\}$ *and* $h_{j,A_j} = n^{-1/2} \sum_{t=1}^{n} \{\tau - G_t(w_{jt})\} x_{jt,A_j}$.

$\overset{A}{\sim}$ *means asymptotically distributed as.*

Proof of Theorem 2

Beginning with the single regressor case, we drop j and k for simplicity. For any vector $\delta \in \mathbb{R}^p$, assume A to be the correct subset model with length p_0 and let

$$Z_n(\delta) = \frac{1}{2} \sum_{t=1}^{n} \{\rho_\tau(u_t - v_t) - \rho_\tau(u_t)\} + \sum_{i=1}^{p} \frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d} \left[\left|\beta_i + \frac{\delta_i}{\sqrt{n}}\right| - |\beta_i|\right], \quad (14)$$

where $u_t = Y_t - x_t^T \beta_0 = \varepsilon_t - w_t$ and $v_t(\delta) = n^{-1/2} x_t^T \delta$.

By taking $Z_n(\delta) = Z_{1n}(\delta) + Z_{2n}(\delta)$, where

$$Z_{1n}(\delta) = \frac{1}{2}\sum_{t=1}^{n}\{\rho_\tau(u_t - v_t) - \rho_\tau(u_t)\}, \tag{15}$$

and

$$Z_{2n}(\delta) = \sum_{i=1}^{p}\frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d}\left[\left|\beta_i + \frac{\delta_i}{\sqrt{n}}\right| - |\beta_i|\right]. \tag{16}$$

Following the steps in [20], the following result could be established by the boundedness of $\{x_t\}$,

$$Z_{1n}(\delta) = \widetilde{Z}_{1n}(\delta) + o_p(1),$$

for each fixed $\delta \in \mathbb{R}^p$, with $\widetilde{Z}_{1n}(\delta) = -\delta^T\zeta_n + \frac{1}{2}\delta^T\Lambda_n\delta$ and $\zeta_n \overset{A}{\sim} \mathcal{N}(h_n, C_n)$.

For Z_{2n}, if $\beta_i \neq 0$, then $|\widehat{\beta}_{\tau,i}|^d \underset{\text{p}}{\to} |\beta_i|^d$ and

$$\sqrt{n}\left[\left|\beta_i + \frac{\delta_i}{\sqrt{n}}\right| - |\beta_i|\right] \to \delta_i sgn(\beta_i).$$

Then, by Slutsky's theorem,

$$\frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d}\left[\left|\beta_i + \frac{\delta_i}{\sqrt{n}}\right| - |\beta_i|\right] \underset{p}{\to} 0.$$

If $\beta_i = 0$, then

$$\frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d}\left[\left|\beta_i + \frac{\delta_i}{\sqrt{n}}\right| - |\beta_i|\right] = |\delta_i|,$$

and

$$\frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d} \underset{p}{\to} \infty.$$

Therefore, we have

$$Z_n(\delta) = \begin{cases} \delta_A^T\zeta_n^A + \frac{1}{2}\delta_A^T\Lambda_A\delta_A & \text{if } \delta_i = 0, \quad \forall i \notin A, \\ \infty & \text{otherwise.} \end{cases}$$

δ_A^T and $\zeta_n^A \in \mathbb{R}^{p_0}$, Λ_A is a $p_0 \times p_0$ matrix. Moreover, under assumption (iii), $\widetilde{Z}_{1n}(\delta)$ has a unique minimiser $\widetilde{\delta}_{1,n} = \Lambda_n^{-1}\zeta_n$ with $\Lambda_n \geq \Lambda_0$ for $n \geq N_0$. Thus, by the convexity of $Z_{1n}(\delta)$, it can be shown that $\widehat{\delta}_{1,n} - \widetilde{\delta}_{1,n} \underset{\text{p}}{\to} 0$, where

$\widehat{\delta}_{1,n}$ denotes the minimiser of $Z_{1n}(\delta)$. Therefore, $\widehat{\delta}_{1,n}$ has the same asymptotic distribution as $\widetilde{\delta}_{1,n}$.

The function $Z_n(\delta)$ could be reparametrised as a function of $\beta = \beta_0 + \delta n^{-1/2}$ as follows

$$Z_n(\delta) = \frac{1}{2}\sum\{\rho_\tau(Y_t - x_t^T\beta) - \rho_\tau(u_t)\} + \sum_{i=1}^{p} \frac{\alpha_n}{\sqrt{n}|\widehat{\beta}_{\tau,i}|^d}\left[\left|\beta_i + \frac{\delta_i}{\sqrt{n}}\right| - |\beta_i|\right].$$

Since this function is minimised at $\beta = \widehat{\beta}^{(AL)}_{\tau,\alpha_n}$, it follows that

$$\widehat{\delta}_n = \frac{1}{\sqrt{n}}(\widehat{\beta}^{(AL)}_{\tau,\alpha_n} - \beta_0).$$

Using the same arguments as in [21] and [22] and assuming $\delta = (\delta_1^T, \delta_2^T)^T$ where δ_1 contains the first p_0 elements of δ, we have

$$\widehat{\delta}_{2,n} \underset{d}{\to} 0,$$

$$\text{and } \widehat{\delta}_{1,n} \underset{d}{\to} \mathcal{N}(\theta_A, \Sigma_A),$$

with $\Sigma_A = \Lambda_A^{-1} C_A \Lambda_A^{-1}$ and $\theta_A = \Lambda_A^{-1} h_{p_0}$.

Now, consider the case of multiple regressors. As pointed out in [20], let $q = 2$. In order to prove the asymptotic normality of $\widehat{\beta}_{1n}$ and $\widehat{\beta}_{2n}$, it is sufficient to consider $Z_n(\delta_1, \delta_2) = Z_n^{(1)}(\delta_1) + Z_n^{(2)}(\delta_2)$, where $Z_n^{(j)}$ is defined in the same way as $Z_n(\delta)$ in equation (14) replacing x_t by the multi-regressor x_{jt}. By taking $Z_{1n}^{(j)}$ and $Z_{2n}^{(j)}$ as defined in equations (15) and (16) respectively, it has been shown that $Z_{1n}^{(j)}(\delta_j) = -\delta_j^T\zeta_{jn} + \frac{1}{2}\delta_j^T\Lambda_{jn}\delta_j + o_p(1)$ uniformly in $\delta_j \in \Delta$, where $\zeta_{jn} \overset{A}{\sim} \mathcal{N}(h_{jn}, C_{jjn})$. Therefore, we only need to calculate $Cov(\zeta_{1n}, \zeta_{2n})$ which is by definition $Cov(\zeta_{1n}, \zeta_{2n}) = \Lambda_{12n}$. The remaining argument is the same as in the one regressor setting.

8.2. Proof of Theorem 1

To establish the results in Theorem 1, it is sufficient to apply the results in Theorem 2 to $Y_t - \lambda$, with $x_{jt} = x_t(\omega_j)$, $\beta_0 = 0$ and $u_t = 0$ such that $\widehat{\beta}_{jn} = \widehat{\beta}(\omega_j)$, $\epsilon_t = Y_t - \lambda$ and $w_{jt} = 0$.

Then, we have $G_t(u) = F_t(\lambda + u)$ and $G_{ts}(u, v) = F_{ts}(\lambda + u, \lambda + v)$. This implies that $g_t(w_{jt}) = f_t(\lambda) = \kappa$ by (C1), so (i) is satisfied. Similarly, (ii) is satisfied because using (C2), we have

$$G_t(u + w_{jt}) - G_t(w_{jt}) = F_t(\lambda + u) - F_t(\lambda) = g_t(w_{jt})u + O(u^{a+1}).$$

Now, let $\{\delta_j\}$ denote the Kronecker delta sequence. Then, we have by the trigonometric identities

$$\frac{1}{n}\sum_{t=1}^{n} x_t(\omega_j)x_t^T(\omega_k) = \frac{1}{2}\delta_{j-k}I + o(1). \tag{17}$$

We note that Λ_{j,A_j} is directly implied from assumption (iv) and the form of θ_{A_j} in Theorem 2. Then, we have from (C2),

$$\Lambda_{j,A_j} = n^{-1}\sum_{t=1}^{n} f_t(\lambda)x_{t,A_j}(\omega_j){x_{t,A_j}}^T(\omega_j) = \frac{1}{2}\kappa I_{p_j} + o(1),$$

where I_{p_j} is an identity matrix with p_j elements, such that (iii) is satisfied. We have

$$x_t x_{t-h}^T = \begin{bmatrix} c_h c_t^2 + s_h c_t s_t & c_h c_t s_t - s_h c_t^2 \\ c_h c_t s_t + s_h s_t^2 & c_h s_t^2 - s_h c_t s_t \end{bmatrix},$$

where $c_t = \cos(\omega t)$ and $s_t = \sin(\omega t)$. It has been shown in [20] that $\sum c_t^2 = \frac{1}{2}n + o(1)$, $\sum s_t^2 = \frac{1}{2}n + o(1)$ and $\sum c_t s_t = O(1)$. Therefore, for any fixed h,

$$n^{-1}\sum_{t=\max(1,1+h)}^{\min(n,n+h)} x_t(\omega_j)x_{t-h}^T(\omega_k) \to M_{j,h} = \frac{1}{2}\begin{bmatrix} c_{j,h} & -s_{j,h} \\ s_{j,h} & c_{j,h} \end{bmatrix}.$$

Then, we have

$$n^{-1}\sum_{t=\max(1,1+h)}^{\min(n,n+h)} x_{t,A_j}(\omega_j){x_{t-h,A_j}}^T(\omega_k) \to \begin{cases} c_{j,h}, \text{ if } p_j = 1, \\ M_{j,h}, \text{ if } p_j = 2. \end{cases}$$

By (C3), we have

$$r_{ts}(w_{jt}, w_{ks}) = F_{ts}(\lambda,\lambda) - F_t(\lambda)F_s(\lambda) = F_{ts}(\lambda,\lambda) - \tau^2 = \tau(1-\tau) - (1/2)\gamma_{t-s}.$$

Now, since $\{r_h\}$ is absolutely summable, we obtain $C_{jn} \to \sum r_h M_{j,h}$. Because r_h is an even function of h, we have $\sum r_h s_h = 0$ and $\sum r_h c_{j,h} = \frac{1}{2}\tau(1-\tau)S(\omega_j)$.

From the above equations and using the results of Theorem 2, C_{jkn} from (iii) reduces to

$$C_{j,A_j} = n^{-1} \sum_{|h|<n} \{\tau(1-\tau) - (1/2)\gamma_h\} \sum_{t=\max(1,1+h)}^{\min(n,n+h)} x_{t,A_j}(\omega_j) x^T_{t-h,A_j}(\omega_k)$$
$$= \frac{1}{2}\tau(1-\tau) S(\omega_j)\delta_{j-k} I_{p_j} + o(1).$$

Hence, (iv) is satisfied. Finally, under (C2), we have

$$h_{jp_j} = n^{-1/2} \sum_{t=1}^{n} \{\tau - F_t(\lambda)\} x_{t,A_j}(\omega_j) = 0.$$

Therefore, we derive

$$\sqrt{n} vec[\widehat{\beta}^{(AL)}_{j,1}(\omega_j)]_{j=1}^{q} \overset{A}{\sim} \mathcal{N}(0, (2/\kappa^2)\tau(1-\tau)\mathbf{S}),$$

where $\mathbf{S} = diag\{S_{p_j}(\omega_1), .., S_{p_j}(\omega_q)\}$, such that $S_{p_j}(\omega_j) = S(\omega_j)$ if $p_j = 1$ and $S_{p_j}(\omega_j) = \{S(\omega_j), S(\omega_j)\}$ if $p_j = 2$. This implies that the $Q^{(AL)}_{\tau,\alpha}(\omega_j)$ are asymptotically independent and

$$\{Q^{(AL)}_{\tau,\alpha}(\omega_j)\} \overset{A}{\sim} \{(1/2)\eta^2 S(\omega_j)(\xi_{p_j})\},$$

for each j with $\eta^2 = \tau(1-\tau)/\kappa^2$ and $\xi_{p_j} \sim \chi^2_{p_j}$ for $p_j \in \{1, 2\}$.

References

[1] T.-H. Li, Quantile periodograms, Journal of the American Statistical Association 107 (498) (2012) 765–776.

[2] P. Bloomfield, Fourier Analysis of Time Series: an Introduction, John Wiley & Sons, 2004.

[3] R. Koenker, G. Bassett Jr, Regression quantiles, Econometrica: Journal of the Econometric Society (1978) 33–50.

[4] A. Hagemann, Robust spectral analysis.

[5] O. Linton, Y.-J. Whang, The quantilogram: With an application to evaluating directional predictability, Journal of Econometrics 141 (1) (2007) 250–282.

[6] H. Dette, M. Hallin, T. Kley, S. Volgushev, et al., Of copulas, quantiles, ranks and spectra: An $l\_\{1\}$-approach to spectral analysis, Bernoulli 21 (2) (2015) 781–831.

[7] J. Lee, S. S. Rao, The quantile spectral density and comparison based tests for nonlinear time series, arXiv preprint arXiv:1112.2759.

[8] Y. Lim, H.-S. Oh, Composite quantile periodogram for spectral analysis, Journal of Time Series Analysis 37 (2) (2016) 195–221.

[9] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) (1996) 267–288.

[10] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, Journal of the American Statistical Association 96 (456) (2001) 1348–1360.

[11] H. Zou, The adaptive lasso and its oracle properties, Journal of the American Statistical Association 101 (476) (2006) 1418–1429.

[12] S. S. Chen, D. L. Donoho, Application of basis pursuit in spectrum estimation, in: Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, Vol. 3, IEEE, 1998, pp. 1865–1868.

[13] T. Kato, M. Uemura, Period analysis using the least absolute shrinkage and selection operator (lasso), Publications of the Astronomical Society of Japan 64 (6) (2012) 122.

[14] P. J. Brockwell, R. A. Davis, Time Series: Theory and Methods, Springer Science & Business Media, 2013.

[15] Y. Wu, Y. Liu, Variable selection in quantile regression, Statistica Sinica (2009) 801–817.

[16] G. Wise, A. Traganitis, J. Thomas, The effect of a memoryless nonlinearity on the spectrum of a random process, IEEE Transactions on Information Theory 23 (1) (1977) 84–89.

[17] H. Wang, C. Leng, Unified lasso estimation by least squares approximation, Journal of the American Statistical Association 102 (479) (2007) 1039–1048.

[18] T.-H. Li, Laplace periodogram for time series analysis, IBM research report (2008) RC24473.

[19] P. Whittle, The simultaneous estimation of a time series harmonic components and covariance structure, Trabajos de estadística 3 (1-2) (1952) 43–57.

[20] T.-H. Li, Laplace periodogram for time series analysis, Journal of the American Statistical Association 103 (482) (2008) 757–768.

[21] K. Knight, et al., Limiting distributions for l_1 regression estimators under general conditions, The Annals of Statistics 26 (2) (1998) 755–770.

[22] R. Koenker, Quantile regression, no. 38, Cambridge university press, 2005.