DETECTION AND ESTIMATION OF HIDDEN PERIODICITY IN ASYMMETRIC NOISE BY USING QUANTILE PERIODOGRAM

Ta-Hsin Li

IBM T. J. Watson Research Center Yorktown Heights, NY 10598-0218, USA thl@us.ibm.com

ABSTRACT

This paper addresses the problem of detecting and estimating hidden periodicity from noisy observations when the noise distribution is asymmetric with heavy tail on one side. The ordinary periodogram is less effective in handling such noise. In this paper, we introduce an alternative periodogram-like function, called the quantile periodogram. The quantile periodogram is constructed from trigonometric regression where a specially designed objective function is used to substitute the squared ℓ_2 norm that leads to the ordinary periodogram. Simulation results are provided to demonstrate the superior performance of the quantile periodogram in comparison with the ordinary periodogram when the noise is asymmetrically distributed with a heavy tail. The asymptotic distribution of the quantile periodogram is derived under the white noise assumption. Extensions to the multivariate case and the complex case are also discussed.

1. INTRODUCTION

Detection and estimation of hidden periodicity with unknown frequency in a noisy data record is traditionally handled by the periodogram. Fisher's test for hidden periodicity, in particular, is based on the maximum of standardized periodogram ordinates [1]. Maximizing the periodogram as a continuous function of the frequency variable often produces very accurate frequency estimates [2] [3]. While powerful under regular conditions such as Gaussian white noise, the periodogram suffers from considerable degradation of performance when the noise has a heavy-tailed distribution.

If the noise can be adequately modeled by a parametric distribution, the maximum likelihood approach usually leads to most effective procedures. But in many cases such models do not exist. In order to accommodate these situations, which is the focus of this paper, one needs robust procedures [4] that perform reasonably well under regular conditions but exhibit greater robustness under heavy tailed conditions. The methods considered in [5]–[7] are such examples. These methods are most effective in situations where the noise distribution is symmetric with heavy tails on both sides, an example being

the Laplace distribution. When the noise has an asymmetric distribution and the heavy tail is only on one side, these methods become less effective.

In this paper, we proposed an alternative robust method suitable for asymmetrically distributed heavy-tail noise. It is based on a periodogram-like function which we call the quantile periodogram. Motivated by quantile regression [8], we construct the quantile periodogram from trigonometric regression by using an alternative criterion to replace the leastsquares criterion that leads to the ordinary periodogram.

In this paper, we demonstrate the effectiveness of the quantile periodogram for detection and estimation of hidden periodicity in asymmetrically distributed heavy-tail noise. We derive the asymptotic distribution of the quantile periodogram under the white noise condition. We also generalize the quantile periodogram to complex-valued observations and to the multivariate case for estimating multiple frequencies.

2. QUANTILE PERIODOGRAM

Given a real-valued time series $\{Y_1, \ldots, Y_n\}$ of length *n*, consider the trigonometric quantile regression problem that minimizes the objective function

$$J_n(\lambda, \boldsymbol{\beta}; \omega) := \sum_{t=1}^n \rho_\tau(Y_t - \lambda - \mathbf{x}_t^T(\omega)\boldsymbol{\beta})$$
(1)

with respect to $\lambda \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^2$, where

$$\mathbf{x}_t(\boldsymbol{\omega}) := [\cos(\boldsymbol{\omega}t), \sin(\boldsymbol{\omega}t)]^T$$

is the trigonometric regressor with frequency ω and $\rho_{\tau}(y)$ is a nonnegative function indexed by $\tau \in (0, 1)$ that takes the form

$$\rho_{\tau}(y) := \begin{cases} \tau y & \text{if } y \ge 0, \\ -(1-\tau)y & \text{if } y < 0. \end{cases}$$

For any fixed ω , let $\hat{\lambda}_n(\omega)$ and $\hat{\boldsymbol{\beta}}_n(\omega)$ denote the minimizer of $J_n(\lambda, \boldsymbol{\beta}; \omega)$. Let $\hat{\lambda}_n$ denote the minimizer of

$$J_n(\lambda) := \sum_{t=1}^n \rho_{\tau}(Y_t - \lambda).$$

The dependence of these minimizers on τ is suppressed in this notation for simplicity. Note that $\hat{\lambda}_n$ is just the τ -th sample quantile of $\{Y_1, \ldots, Y_n\}$ [8]. Note also that $\hat{\lambda}_n(\omega)$ and $\hat{\boldsymbol{\beta}}_n(\omega)$ can be obtained by linear programming techniques [8].

With this notation, let the $\tau\text{-th}$ quantile periodogram be defined as

$$Q_n(\boldsymbol{\omega}) := J_n(\hat{\lambda}_n) - J_n(\hat{\lambda}_n(\boldsymbol{\omega}), \hat{\boldsymbol{\beta}}_n(\boldsymbol{\omega}); \boldsymbol{\omega}).$$
(2)

For any fixed ω , it is easy to see that $Q_n(\omega)$ coincides with the generalized log likelihood ratio statistic for testing the null hypothesis $H_0: \beta = 0$ against the alternative hypothesis $H_1: \beta \neq 0$ under the assumption that

$$Y_t = \lambda + \mathbf{x}_t^T(\boldsymbol{\omega})\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \qquad (t = 1, \dots, n),$$

where $\{\varepsilon_t\}$ is an i.i.d. sequence with probability density function $f_{\varepsilon}(y) \propto \exp\{-\rho_{\tau}(y)\}$ which is known as an asymmetric Laplace distribution.

The quantile periodogram $Q_n(\omega)$ is closely related to the ordinary periodogram which is usually defined as

$$I_n(\omega) := n^{-1} \left| \sum_{t=1}^n Y_t \exp(-it\omega) \right|^2$$

Indeed, it is easy to verify that for any Fourier frequency ω (i.e., an integral multiple of $2\pi/n$) the ordinary periodogram can be expressed as

$$I_n(\boldsymbol{\omega}) = L_n(\tilde{\lambda}_n) - L_n(\tilde{\lambda}_n(\boldsymbol{\omega}), \tilde{\boldsymbol{\beta}}_n(\boldsymbol{\omega}); \boldsymbol{\omega}),$$

where $\tilde{\lambda}_n(\omega)$ and $\tilde{\boldsymbol{\beta}}_n(\omega)$ denote the minimizer of

$$L_n(\lambda,\boldsymbol{\beta};\boldsymbol{\omega}) := \frac{1}{2} \sum_{t=1}^n |Y_t - \lambda - \mathbf{x}_t^T(\boldsymbol{\omega})\boldsymbol{\beta}|^2$$

and λ_n denotes the minimizer of

$$L_n(\lambda) := \frac{1}{2} \sum_{t=1}^n |Y_t - \lambda|^2.$$

Note that λ_n is just the sample mean of $\{Y_1, \ldots, Y_n\}$. As we can see, the quantile periodogram $Q_n(\omega)$ is constructed by simply replacing the least-squares criterion with the quantile regression criterion.

3. DETECTION OF HIDDEN PERIODICITY

To detect a hidden periodicity with unknown frequency and unknown noise distribution, a traditional method is Fisher's test [1] which employs the statistic

$$g := \frac{\max\{I_n(\omega_k)\}}{\sum I_n(\omega_k)}$$

where the ω_k are the Fourier frequencies in the interval $(0, \pi)$. Under the white noise assumption, the periodogram ordinates



Fig. 1. ROC curves for detection of hidden periodicity using the *g* and g_Q statistics. Solid line, the g_Q test with $\tau = 0.4$; dashed line, the g_Q test with $\tau = 0.5$; dotted line, the g_Q test with $\tau = 0.6$; dash-dotted line, the *g* test (Fisher). Results are based on 5000 Monte Carlo runs.

 $I_n(\omega_k)$ are approximately distributed as i.i.d. $(1/2)\sigma^2\chi_2^2$ random variables, where σ^2 is the variance of the noise. As a result, the distribution of *g* under H_0 can be approximated by a well-known distribution so that a suitable threshold can be easily selected for the test [1, pp. 91–96].

A similar statistic can be constructed on the basis of the quantile periodogram. Indeed, let

$$g_Q := \frac{\max\{Q_n(\omega_k)\}}{\sum Q_n(\omega_k)}$$

The following theorem shows that for large sample sizes the quantile periodogram ordinates $Q_n(\omega_k)$ behave similarly to the ordinary periodogram ordinates $I_n(\omega_k)$, so the justification for Fisher's test which is based on *g* remains valid for the new test which is based on g_Q .

Theorem 1. [9] Let $\{Y_t\}$ be an i.i.d. sequence with probability distribution function F(y) satisfying $F(y + \lambda) - F(\lambda) = \dot{F}(\lambda)y + O(|y|^2)$ for small |y| and $\dot{F}(\lambda) > 0$, where λ denotes the τ -th quantile of F(y) such that $F(\lambda) = \tau$. Let $\{\omega_k\}$ be a finite set of Fourier frequencies. Then, as $n \to \infty$, $\{Q_n(\omega_k)\}$ is asymptotically distributed as $\{(1/2)\eta^2\xi_k\}$, where the ξ_k are i.i.d. χ_2^2 random variables and $\eta^2 := \tau(1 - \tau)/\dot{F}(\lambda)$.

To compare the statistical performance of these tests, Fig. 1 depicts their ROC curves based on a Monte Carlo simulation. In this experiment, the time series takes the form

$$Y_t = \cos(\omega_0 t + \phi) + e_t$$
 $(t = 1, ..., n),$ (3)

where $\omega_0 = 2\pi \times 0.1$, $\phi \sim U(-\pi, \pi)$, and n = 50. The noise $\{e_t\}$ takes the form $e_t = c(\zeta_t - \mu)$, where the ζ_t are i.i.d. lognormal random variables such that $\{\log(\zeta_t)\} \sim \text{i.i.d. } N(0,1)$. The location and scale parameters μ and c > 0 are chosen such that $\{e_t\}$ has mean zero and the SNR equals -6 dB. As we can see from Fig. 1, the g_Q statistic outperforms the gstatistic by a large margin. Note that the g_Q statistic performs better with $\tau = 0.4$ than it does with $\tau = 0.5$ or 0.6 because the effective SNR in the quantile periodogram [6] [9], which is proportional to $1/\eta^2$, is higher when $\tau = 0.4$.

To better understand this behavior, Fig. 2 depicts the ordinary periodogram and the quantile periodogram (with $\tau = 0.4$ and 0.5) for a realization of the time series under H_1 . In this example, the ordinary periodogram completely misses the hidden periodicity whereas the quantile periodogram detects it with a large spike at the correct frequency. Fig. 2(a) shows that strong oscillation exists in the time series plot not around the sample mean but around a much lower level. This is due entirely to the fact that the noise has a much heavier upper tail. What the ordinary periodogram depicts is essentially the oscillations around the mean. The quantile periodogram, on the other hand, is capable of depicting oscillations at other levels with a suitable choice of τ .

4. FREQUENCY ESTIMATION

As with the ordinary periodogram [2], the unknown frequency of the hidden periodicity can be estimated by the maximizer of a quantile periodogram.

To demonstrate, let $\{Y_t\}$ take the form (3) as in the previous section except that the frequency parameter ω_0 is also randomized to have a uniform distribution between $2\pi \times 0.1$ and $2\pi \times 0.1 + 2\pi/n$. The maximizer of the quantile periodogram as a continuous function of ω is obtained numerically with the function optimize in the software package R which implements a bisection algorithm. The initial interval for the bisection search is taken to be $(\tilde{\omega}_n - 2\pi/n, \tilde{\omega}_n + 2\pi/n)$, where $\tilde{\omega}_n$ denotes the maximizer of the quantile periodogram constrained at the Fourier frequencies. The maximizer of the ordinary periodogram is computed in the same way.

Fig. 3 depicts the mean-square error (MSE) for estimating $f_0 := \omega_0/(2\pi)$ by maximizing the quantile and ordinary periodograms for different values of SNR. Note that the maximizer of the quantile periodogram with $\tau = 0.5$ is nothing but the nonlinear least absolute deviations (NLAD) estimator discussed in [7], which is known to be robust against symmetrically distributed heavy-tailed noise.

As can be seen from Fig. 3, while the NLAD estimator remains superior to the ordinary periodogram maximizer in this case, the quantile periodogram maximizer with the choice $\tau = 0.4$ produces even more accurate estimates across the SNR range considered. The quantile periodogram maximizer also reduces the SNR threshold, a value of the SNR below which the estimation accuracy deteriorates rapidly.



Fig. 2. A time series with hidden periodicity and its ordinary and quantile periodograms as functions of $f := \omega/(2\pi)$. (a) The time series. (b) Ordinary periodogram. (c) Quantile periodogram with $\tau = 0.4$. (d) Quantile periodogram with $\tau = 0.5$. Dotted line in (a) depicts the sample mean. Dashed line in (b)–(d) shows the location of the hidden frequency. All periodograms are standardized so as to add up to unity.



Fig. 3. Mean-square error for frequency estimation by maximizing the ordinary and quantile periodograms. Slid line, quantile periodogram with $\tau = 0.4$; dashed line, quantile periodogram with $\tau = 0.5$ (NLAD); dotted line, quantile periodogram with $\tau = 0.6$; dash-dotted line, ordinary periodogram. Results are based on 5000 Monte Carlo runs.

5. EXTENSIONS

The methodology can be easily extended to the case of multiple periodicities by considering the objective function

$$J_n(\lambda,\boldsymbol{\beta};\boldsymbol{\omega}) := \sum_{t=1}^n \rho_\tau(Y_t - \lambda - \mathbf{x}_t^T(\boldsymbol{\omega})\boldsymbol{\beta}),$$

where $\boldsymbol{\omega} := [\omega_1, \dots, \omega_p]^T$ is the multivariate frequency variable, $\mathbf{x}_t(\boldsymbol{\omega}) := [\cos(\omega_1 t), \sin(\omega_1 t), \dots, \cos(\omega_p t), \sin(\omega_p t)]$ is the corresponding trigonometric regressor. Let $\hat{\lambda}_n(\boldsymbol{\omega})$ and $\hat{\boldsymbol{\beta}}_n(\boldsymbol{\omega})$ denote the minimizer of $J_n(\lambda, \boldsymbol{\beta}; \boldsymbol{\omega})$ for fixed $\boldsymbol{\omega}$. Then, the multivariate τ -th quantile periodogram can be defined as

$$Q_n(\boldsymbol{\omega}) := J_n(\hat{\lambda}_n) - J_n(\hat{\lambda}_n(\boldsymbol{\omega}), \hat{\boldsymbol{\beta}}_n(\boldsymbol{\omega}); \boldsymbol{\omega}),$$

where $\hat{\lambda}_n$ and $J_n(\lambda)$ are the same as in the univariate case. As a generalization of the multivariate Laplace periodogram discussed in [10], the multivariate quantile periodogram is capable of dealing with closely spaced hidden periodicities.

The methodology can also be extended to the complex case by replacing $\rho_{\tau}(y)$ with

$$\rho_{\tau,c}(\mathbf{y}) := 2\{\rho_{\tau}(\mathfrak{R}(\mathbf{y})) + \rho_{\tau}(\mathfrak{I}(\mathbf{y}))\}$$

and considering the objective function

$$J_{n,c}(\lambda,\boldsymbol{\beta};\boldsymbol{\omega}) := \sum_{t=1}^{n} \rho_{\tau,c}(Y_t - \lambda - \mathbf{x}_t^T(\boldsymbol{\omega})\boldsymbol{\beta}),$$

where $\mathbf{x}_t(\boldsymbol{\omega}) := [\exp(i\omega_1 t), \dots, \exp(i\omega_p t)]^T$ is the complex trigonometric regressor and where $\lambda \in \mathbb{C}$ and $\boldsymbol{\beta} \in \mathbb{C}^p$ are complex parameters. With $\hat{\lambda}_n(\boldsymbol{\omega})$ and $\hat{\boldsymbol{\beta}}_n(\boldsymbol{\omega})$ denoting the minimizer of this objective function for fixed $\boldsymbol{\omega}$, and with $\hat{\lambda}_n$ denoting the minimizer of

$$J_{n,c}(\lambda) := \sum_{t=1}^n \rho_{\tau,c}(Y_t - \lambda),$$

the corresponding quantile periodogram can be defined as

$$Q_{n,c}(\boldsymbol{\omega}) := J_{n,c}(\hat{\lambda}_n) - J_{n,c}(\hat{\lambda}_n(\boldsymbol{\omega}), \hat{\boldsymbol{\beta}}_n(\boldsymbol{\omega}); \boldsymbol{\omega}).$$

In the univariate case, an asymptotic theory similar to Theorem 1 can be developed under the assumption that the real and imaginary parts of $\{Y_t\}$ are i.i.d. sequences with probability distribution function F(y) which satisfies the assumptions of Theorem 1. The only difference is the scaling constant takes the form $\eta_c^2 := 2\eta^2$. This is consistent with the doubling of the power spectrum for complex white noise in comparison with the power spectrum of the real and imaginary parts of the complex white noise.

6. REFERENCES

- P. Bloomfield, Fourier Analysis of Time Series, 2nd Edition, New York, Wiley, 2000.
- [2] A. M. Walker, "On the estimation of a harmonic component in a time series with stationary independent residuals," *Biometrika*, vol. 58, no. 1, pp. 21–36, 1971.
- [3] S. M. Kay, *Modern Spectral Analysis*, Englewood Cliffs, NJ: Prentice Hall, 1988.
- [4] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, 2nd Edition, New York: Wiley, 2009.
- [5] V. Katkovnik, "Robust M-estimates of the frequency and amplitude of a complex-valued harmonic," *Signal Process.*, vol. 77, no. 1, pp. 71–84, 1999.
- [6] T. H. Li, "Laplace periodogram for time series analysis," J. Amer. Statist. Assoc., vol. 103, no. 482, pp. 757–768, 2008.
- [7] T. H. Li and K. S. Song, "Estimation of the parameters of sinusoidal signals in non-Gaussian noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 62–72, 2009.
- [8] R. Koenker, *Quantile Regression*, Cambridge, UK: Cambridge University Press, 2005.
- [9] T. H. Li, "Quantile periodograms," IBM Research Report RC25199, 2011. Available at domino.research.ibm.com/library/cyberdig.nsf.
- [10] T. H. Li, "A robust periodogram for high-resolution spectral analysis," *Signal Process.*, vol. 90, no. 7, pp. 2133–2140, 2010.