RC24473 (W0801-089) January 22, 2008
Mathematics

IBM Research Report

Laplace Periodogram for Time Series Analysis

Ta-Hsin Li
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

Laplace Periodogram for Time Series Analysis

(Complete Version)

Ta-Hsin Li

Department of Mathematical Sciences

IBM T. J. Watson Research Center

Yorktown Heights, NY 10598-0218

thl@us.ibm.com

November 1, 2007

Abstract

A new type of periodogram, called the Laplace periodogram, is derived by replacing least squares with least absolute deviations in the harmonic regression procedure that produces the ordinary periodogram of a time series. An asymptotic analysis reveals a connection between the Laplace periodogram and the zero-crossing spectrum. This relationship provides a theoretical justification for the Laplace periodogram to be used as a nonparametric tool for analyzing the serial dependence of time-series data. Superiority of the Laplace periodogram in handling heavy-tailed noise and nonlinear distortion is demonstrated by simulations. A real-data example shows its great effectiveness for analyzing heart rate variability in the presence of ectopic events and artifacts.

Abbreviated Title: Laplace periodogram

Key Words and Phrases: frequency, harmonic, heart rate variability, heavy tail, least absolute deviation, mixed spectrum, robust, spectral analysis, zero crossing

1. INTRODUCTION

Regression using least absolute deviations (LAD) is a well-known method of data analysis that can be dated back to more than two centuries ago. With the advance of computing techniques, especially fast linear programming algorithms, LAD regression has regained the attention of both theoretical and applied researchers in modern times. The recent book by Koenker (2005) provides an excellent account of the history and the latest development of LAD regression and its generalization to quantile regression.

The LAD methodology is known for its robustness against outliers. Statistical properties of the LAD estimators have been studied intensively in the context of linear and nonlinear regression (Bloomfield and Steiger 1983; Breidt, Davis, and Trindate 2001; Dodge 1997, 2002; Dielman 2005; Koenker 2005; Lai and Lee 2005). In this article, we consider an application of LAD in the field of time series analysis. Specifically, we propose an LAD-based approach to analyzing the serial dependence of time-series data by replacing least squares (LS) with LAD in the harmonic regression procedure that produces the ordinary periodogram. This leads to a periodogram-like function which we refer to as the Laplace periodogram.

We show that the asymptotic distribution of the Laplace periodogram is directly related to what we call the zero-crossing spectrum in the same way as the ordinary periodogram is related to the autocorrelation spectrum. Zero crossings are known to contain rich information about the dependence characteristics of time series and are widely used in applications such as signal processing (Kedem 1994). The connection with the zero-crossing spectrum provides the theoretical foundation and justification for the Laplace periodogram to be used as a

general-purpose nonparametric tool of time series analysis. In addition, we present some simulation results to demonstrate the superior robustness of the Laplace periodogram in handling heavy-tailed noise and nonlinear distortion. We also discuss its usefulness for time series of mixed spectrum. Finally, we provide a real-data example of heart rate variability analysis where the effectiveness of the Laplace periodogram in dealing with the contamination from ectopic events and artifacts is demonstrated.

2. LAPLACE PERIODOGRAM

For a given time series $\{y_1,\ldots,y_n\}$ of length n and a frequency parameter $\omega \in (0,\pi)$, the ordinary periodogram is defined as

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

where $i := \sqrt{-1}$. It is easy to show that if ω is a Fourier frequency, i.e., if $\omega = 2\pi k/n$ for some integer k, then the ordinary periodogram can also be written as

$$G_n(\omega) = \tfrac{1}{4} n \|\tilde{\boldsymbol{\beta}}_n(\omega)\|^2, \tag{1}$$

where $\tilde{\boldsymbol{\beta}}_n(\omega)$ is the LS regression solution

$$\tilde{\boldsymbol{\beta}}_n(\omega) := \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^2} \sum_{t=1}^{n} |y_t - \boldsymbol{x}_t^T(\omega)\boldsymbol{\beta}|^2 \tag{2}$$

with the harmonic regressor $\boldsymbol{x}_t(\omega) := [\cos(\omega t), \sin(\omega t)]^T$. Since $\tilde{\boldsymbol{\beta}}_n(\omega)$ comes from LS regression, the ordinary periodogram can also be called the Gauss periodogram, which explains why we use the letter G in our notation for the ordinary periodogram rather than the more conventional I (Brockwell and Davis 1991).

Naturally, we would like to replace the LS criterion with the LAD criterion in the harmonic regression procedure (2) to obtain a new regression coefficient

$$\hat{\boldsymbol{\beta}}_n(\omega) := \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^2} \sum_{t=1}^{n} |y_t - \boldsymbol{x}_t^T(\omega)\boldsymbol{\beta}|. \tag{3}$$

Based on the LAD coefficient, we define, in the same way as (1), a new periodogram

$$L_n(\omega) := \tfrac{1}{4} n \|\hat{\boldsymbol{\beta}}_n(\omega)\|^2. \tag{4}$$

We call this function of ω the Laplace periodogram.

As can be seen, both ordinary and Laplace periodograms are proportional to the squared norm of harmonic regression coefficients; the difference is that the former comes from LS regression and the latter from LAD regression. It is expected that the Laplace periodogram will inherit all the robustness benefits of linear LAD regression as discussed in Bloomfield and Steiger (1983) and Koenker (2005).

3. ASYMPTOTIC THEORY

This section contains some theoretical results about the asymptotic distribution of the Laplace periodogram for time series of continuous and mixed spectrum.

3.1 A General Theorem

First, we provide a general result concerning the asymptotic distribution of the LAD regression coefficients. This type of results can be established under various conditions (Arcones 2001; Bantli and Hallin 1999; Bloomfield and Steiger 1983; Koenker 2005; Lai and Lee 2005; Pollard 1991; Portnoy 1991; Weiss 1990; Wu 2007). Our focus is on the situation

where (a) the regressors, denoted by $\{\boldsymbol{x}_{jt}\}$ $(j = 1,\ldots,q)$, are uniformly bounded sequences of deterministic vectors in $\mathbb{R}^p$ for some $p \in \mathbb{N} := \{1,2,\ldots\}$, and may depend on n (not explicitly denoted for simplicity); (b) $\{\varepsilon_t\}$ is a dependent random process; and (c) the regression functions $\boldsymbol{x}_{jt}^T\boldsymbol{\beta}_0$ may not be correctly specified for y_t. Note that the boundedness of the regressors considerably simplifies the conditions.

Theorem 1. *Let* $y_t = \mu_t + \varepsilon_t$ $(t = 1,\ldots,n)$ *and*

$$\hat{\boldsymbol{\beta}}_{jn} := \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{t=1}^{n} |y_t - \boldsymbol{x}_{jt}^T\boldsymbol{\beta}| \quad (j = 1,\ldots,q), \tag{5}$$

where $\{\mu_t\}$ *is a deterministic sequence and* $\{\varepsilon_t\}$ *is an m-dependent random process having marginal distribution functions* $F_t(x)$*, marginal densities* $f_t(x) := \dot{F}_t(x)$*, and bivariate distribution functions* $F_{ts}(u,v)$*. Let* $r_{ts}(u,v) := F_{ts}(u,v) - F_t(u)F_s(v)$*. For any* $\boldsymbol{\beta}_0 \in \mathbb{R}^p$*, let* $w_{jt} := \boldsymbol{x}_{jt}^T\boldsymbol{\beta}_0 - \mu_t$*. Assume that there are positive numbers* d*,* c*, and* N_0 *and positive-definite matrices* $\boldsymbol{W}_0$ *and* $\boldsymbol{Q}_0$ *such that*

(i) $f_t(w_{jt}) = \mathcal{O}(1)$ *uniformly in* $t \in \mathbb{N}$ *for all* j*,*

(ii) $F_t(u + w_{jt}) - F_t(w_{jt}) = f_t(w_{jt})u + \mathcal{O}(u^{d+1})$ *uniformly in* $t \in \mathbb{N}$ *for* $|u| \le c$ *and all* j*,*

(iii) $\boldsymbol{W}_{jkn} := n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} r_{ts}(w_{jt},w_{ks})\boldsymbol{x}_{jt}\boldsymbol{x}_{ks}^T \ge \boldsymbol{W}_0$ *for all* j*,* k*, and* $n \ge N_0$*,*

(iv) $\boldsymbol{Q}_{jn} := n^{-1}\sum_{t=1}^{n} f_t(w_{jt})\boldsymbol{x}_{jt}\boldsymbol{x}_{jt}^T \ge \boldsymbol{Q}_0$ *for all* j *and* $n \ge N_0$*.*

Then, as $n \to \infty$*, we have* $\mathrm{vec}\{\hat{\boldsymbol{\beta}}_{jn} - \boldsymbol{\beta}_0\}_{j=1}^{q} \overset{A}{\sim} N(\boldsymbol{\theta}_n, \boldsymbol{\Sigma}_n)$*, where* $\boldsymbol{\theta}_n := \mathrm{vec}\{\boldsymbol{Q}_{jn}^{-1}\boldsymbol{h}_{jn}\}_{j=1}^{q}$*,* $\boldsymbol{\Sigma}_n := [\boldsymbol{Q}_{jn}^{-1}\boldsymbol{W}_{jkn}\boldsymbol{Q}_{kn}^{-1}]_{j,k=1}^{q}$*,* $\boldsymbol{h}_{jn} := n^{-1/2}\sum_{t=1}^{n}\{\frac{1}{2} - F_t(w_{jt})\}\boldsymbol{x}_{jt}$*, and* $\overset{A}{\sim}$ *means "asymptotically distributed as."*

Proof. See Appendix. □

Remark 1. Under the exact regression model $y_t = \boldsymbol{x}_t^T \boldsymbol{\beta}_0 + \varepsilon_t$ and the assumption that $\{\varepsilon_t\}$ is an ARMA process, Davis and Dunsmuir (1997) obtained an asymptotic normality result for the LAD estimates of $\boldsymbol{\beta}_0$ and the ARMA parameters. Under somewhat stronger conditions, Portnoy (1991) obtained a similar result for a broader class of dependent processes in the single regressor case. See also Weiss (1990) and Wu (2007).

As we can see, Theorem 1 not only establishes the asymptotic normality of the LAD regression coefficients, but also reveals the impact of model misspecification on the asymptotic distribution. It shows that a model misspecification, manifested as a nonzero 'bias' $w_t = \boldsymbol{x}_t^T \boldsymbol{\beta}_0 - \mu_t$, has impact on the asymptotic mean as well as the asymptotic covariance matrix. This is different from LS regression where the bias only affects the asymptotic mean.

From Theorem 1 and Remark 4 in Appendix, we obtain the following result.

Corollary 1 (Linear Processes)**.** *Let $\{\varepsilon_t\}$ be a linear process of the form $\varepsilon_t = \sum_{j=-\infty}^{\infty} \phi_j e_{t-j}$, where $\{e_t\}$ is an i.i.d. random sequence with $E(|e_t|) < \infty$ and $\{\phi_j\}$ is an absolutely-summable deterministic sequence satisfying $\sum_{|j|>m} |\phi_j| = o(n^{-1})$ for some $m = O(n^{\delta})$ and $\delta \in [0, \frac{1}{4})$. Then, the assertion in Theorem 1 remains true under Assumptions (i)–(iv).*

Proof. See Appendix. □

In the special case where the ε_t's are i.i.d. and the median of y_t is a linear function of $\boldsymbol{x}_t$, the following result can be obtained from Theorem 1. Similar results are given by Bantli and Hallin (1999), Koenker (2005, Theorem 4.1), and Pollard (1991).

Corollary 2 (White Noise)**.** *Let $\hat{\boldsymbol{\beta}}_n$ be defined by (5) with a single regressor $\boldsymbol{x}_t$ and with $y_t = \boldsymbol{x}_t^T \boldsymbol{\beta}_0 + \varepsilon_t$ $(t = 1, \ldots, n)$, where $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with distribution*

function $F(x)$ and density $f(x)$ satisfying $F(0) = \frac{1}{2}$ and $f(0) > 0$. Assume further that

(v) $f(x)$ is continuously differentiable in a neighborhood of $x = 0$,

(vi) $\boldsymbol{D}_n := n^{-1}\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^T \geq \boldsymbol{D}_0 > 0$ for all $n \geq N_0$.

Then, as $n \to \infty$, $n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \overset{A}{\sim} N(\boldsymbol{0}, \eta^2 \boldsymbol{D}_n^{-1})$, where $\eta^2 := 1/(4f^2(0))$.

Proof. The assumptions in Theorem 1 are satisfied with $w_t = 0$, $\boldsymbol{W}_n = \frac{1}{4}\boldsymbol{D}_n$, and $\boldsymbol{Q}_n = f(0)\boldsymbol{D}_n$. Note: the subscripts j and k are droped in the single regressor case. □

A comparison of Corollary 2 with Theorem 1 reveals that the serial dependence of $\{\varepsilon_t\}$ manifests itself entirely through $r_{ts}(u,v)$ in the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_n$. Therefore, it is worthwhile to take a closer look at this quantity.

First, since $F_{tt}(u,v) = F_t(\min(u,v))$, we have $r_{tt}(u,v) = F_t(\min(u,v)) - F_t(u)F_t(v)$, and in particular, $r_{tt}(u,u) = F_t(u)\{1 - F_t(u)\}$. If the ε_t's are independent, then $F_{ts}(u,v) = F_t(u)F_s(v)$ for $t \neq s$, so that $r_{ts}(u,v) \equiv 0$ for $t \neq s$. Moreover, $r_{ts}(w_{jt}, w_{ks})$ is nothing but the cross covariance between the binary (0-1) random processes $\{I(\varepsilon_t < w_{jt})\}$ and $\{I(\varepsilon_t < w_{kt})\}$. In particular, $r_{ts}(0,0) = \mathrm{Cov}\{I(\varepsilon_t < 0), I(\varepsilon_s < 0)\}$.

Generally, for any fixed u and v, let us define $\gamma_{ts}(u,v) := P\{\varepsilon_t < u, \varepsilon_s > v\} + P\{\varepsilon_t > u, \varepsilon_s < v\}$ as the *joint level-crossing rate* of $\{\varepsilon_t\}$ at t and s. Under the assumption that $\{\varepsilon_t\}$ has continuous univariate and bivariate distributions, we can write $\gamma_{ts}(u,v) = F_t(u) + F_s(v) - 2F_{ts}(u,v)$. This implies that $F_{ts}(u,v) = \frac{1}{2}\{F_t(u) + F_s(v) - \gamma_{ts}(u,v)\}$, so $r_{ts}(u,v) = \frac{1}{2}\{F_t(u) + F_s(v) - 2F_t(u)F_s(v)\} - \frac{1}{2}\gamma_{ts}(u,v)$. In the special case of $F_t(0) = \frac{1}{2}$ for all t, we have $r_{ts} := r_{ts}(0,0) = \frac{1}{4} - \frac{1}{2}\gamma_{ts}$, where $\gamma_{ts} := \gamma_{ts}(0,0) = P\{\varepsilon_t\varepsilon_s < 0\} = \frac{1}{2}(1 - 4r_{ts})$ is the probability that there are odd number of zero crossings between t and s. We call γ_{ts} the *lagged zero-crossing*

rate of $\{\varepsilon_t\}$ between t and s. If $F_t(0) = \frac{1}{2}$ for all t, we can write $\gamma_{ts} = 1 - 2p_{ts}$, where $p_{ts} := F_{ts}(0,0) = P\{\varepsilon_t < 0, \varepsilon_s < 0\}$ is known as an orthant probability.

Based on the general results presented in Theorem 1, Corollary 1, and Corollary 2, an asymptotic theory can be developed for the Laplace periodogram with the choice of $\boldsymbol{x}_{jt} := \boldsymbol{x}_t(\omega_j)$ for some $\omega_j \in (0,\pi)$. In the remainder of this section, we consider two types of time series: those of continuous spectrum and those of mixed spectrum. We begin with time series of continuous spectrum.

3.2 Asymptotics for Time Series of Continuous Spectrum

The simplest time series of continuous spectrum is an i.i.d. white noise process. Our first theorem focuses on this type of time series.

Theorem 2 (White Noise)**.** *Let $\hat{\boldsymbol{\beta}}_n(\omega)$ and $L_n(\omega)$ be defined by (3) and (4) with $y_t = \varepsilon_t$ $(t = 1,\ldots,n)$, where $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with distribution function $F(x)$ and density $f(x)$ such that $F(0) = \frac{1}{2}$ and $f(0) > 0$. Assume that $f(x)$ also satisfies Assumption (v) in Corollary 2. Let $\{\omega_1,\ldots,\omega_q\}$ be a set of distinct values in $(0,\pi)$ which may depend on n but which satisfy the condition*

(vii) $D_{jkn} := n^{-1}\sum_{t=1}^{n} \boldsymbol{x}_t(\omega_j)\boldsymbol{x}_t^T(\omega_k) = \frac{1}{2}\delta_{j-k}\mathbf{I} + \mathcal{O}(1)$,

where $\{\delta_j\}$ is the Kronecker delta sequence such that $\delta_0 = 1$ and $\delta_j = 0$ for $j \neq 0$. Then, as $n \to \infty$, we have $n^{1/2}\mathrm{vec}\{\hat{\boldsymbol{\beta}}_n(\omega_j)\}_{j=1}^{q} \overset{A}{\sim} N(\mathbf{0}, 2\eta^2\mathbf{I})$ and $\{L_n(\omega_j)\} \overset{A}{\sim} \{\frac{1}{2}\eta^2 Z_j\}$, where $\eta^2 := 1/(4f^2(0))$ and $Z_j \sim$ i.i.d. $\chi^2(2)$ $(j = 1,\ldots,q)$.

Proof. See Appendix. □

Remark 2. Assumption (vii) is satisfied by any distinct ω_j's that do not depend on n. It is also satisfied by the Fourier frequencies. In general, certain separation conditions must hold for the ω_j's to satisfy Assumption (vii).

According to Theorem 2, the asymptotic mean of $L_n(\omega)$ is equal to η^2 for all ω. To compare this expression with the corresponding result of the ordinary periodogram, let us recall that if the noise has a finite variance σ^2, then it can be shown (Brockwell and Davis 1991) that $G_n(\omega) \overset{A}{\sim} \frac{1}{2}\sigma^2\chi^2(2)$, so the asymptotic mean of $G_n(\omega)$ is equal to σ^2. As can be seen, η^2 plays the role of σ^2. For Gaussian distributions, $\eta^2 = \frac{1}{2}\pi\sigma^2$, and for double exponential (Laplace) distributions, $\eta^2 = \frac{1}{2}\sigma^2$. In general, a finite η^2 does not require a finite variance. For example, for Cauchy noise with scale parameter $\sigma > 0$, $\eta^2 = \frac{1}{4}\pi^2\sigma^2$ is finite but the variance (and the mean) equals infinity. This is an exemplary situation where the Laplace periodogram out-shines the ordinary periodogram: the Laplace periodogram does not require the existence of any moments to have a well-defined asymptotic distribution. It means in practical terms that the Laplace periodogram is more robust to high volatilities in the data.

Now we consider the more general situation of dependent processes. To do so, we need to introduce a new concept of stationarity.

Definition 1 (Stationarity in Zero Crossings)**.** The lagged zero-crossing rate of a random process $\{\varepsilon_t\}$ between t and s is defined as $\gamma_{ts} := P\{\varepsilon_t\varepsilon_s < 0\}$ and $\{\varepsilon_t\}$ is said to be stationary in zero crossings if and only if γ_{ts} depends only on $t-s$, i.e., $\gamma_{ts} = \gamma_{t-s}$, for all t and s. In this case, we refer to γ_τ as the lag-τ zero-crossing rate of $\{\varepsilon_t\}$ and refer to $S(\omega) := \sum_{\tau=-\infty}^{\infty}(1-2\gamma_\tau)\cos(\omega\tau)$ as the zero-crossing spectrum of $\{\varepsilon_t\}$.

Stationarity in zero crossings can be characterized in a number of ways. The following lemma is just an example.

Lemma 1. *Let $\{\varepsilon_t\}$ be a random process with marginal distribution functions $F_t(x)$ satisfying $F_t(0) = \frac{1}{2}$ for all t. Then, the following statements are equivalent.*

(a) the process $\{\varepsilon_t\}$ is stationary in zero crossings with zero-crossing rates γ_τ;

(b) the orthant probabilities $p_{ts} := P\{\varepsilon_t < 0, \varepsilon_s < 0\}$ depend only on $t-s$, i.e., $p_{ts} = p_{t-s}$;

(c) the process $\{I(\varepsilon_t < 0)\}$ is stationary in second moments with autocovariances r_τ.

In this case, $\gamma_\tau = 1 - 2p_\tau = \frac{1}{2}(1 - 4r_\tau)$.

Remark 3. A strictly stationary process, such as a linear process considered in Corollary 1, is stationary in zero crossings. Since γ_τ is even and $\gamma_0 = 0$, we can write $S(\omega) = 2\sum_{\tau=0}^{\infty}(1 - 2\gamma_\tau)\cos(\omega\tau) - 1$. For white noise, $r_\tau = \frac{1}{4}\delta_\tau$, $\gamma_\tau = \frac{1}{2}(1-\delta_\tau)$, $p_\tau = \frac{1}{4}(1+\delta_\tau)$, and $S(\omega) = 1$.

Equipped with the concept of stationarity in zero crossings, the following result can be obtained from Theorem 1.

Theorem 3 (Dependent Processes)**.** *Let $\hat{\boldsymbol{\beta}}_n(\omega)$ and $L_n(\omega)$ be defined by (3) and (4) with $y_t = \varepsilon_t$ $(t = 1, \ldots, n)$, where $\{\varepsilon_t\}$ is (a) an m-dependent process stationary in zero crossings or (b) a linear process defined in Corollary 1, in either case having zero-crossing rates $\{\gamma_\tau\}$ such that $\sum_{\tau=0}^{\infty}|1 - 2\gamma_\tau| < \infty$ and having a common marginal distribution function $F(x)$ and density $f(x)$ such that $F(0) = \frac{1}{2}$ and $f(0) > 0$ and that Assumption (v) in Corollary 2 is satisfied. Let $\{\omega_1, \ldots, \omega_q\}$ be a set of distinct values in $(0, \pi)$ satisfying Assumption (vii) in Theorem 2. Assume that $S(\omega_j) > 0$ for all j. Then, as $n \to \infty$, $n^{1/2}\,\mathrm{vec}\{\hat{\boldsymbol{\beta}}_n(\omega_j)\}_{j=1}^{q} \overset{A}{\sim} N(\mathbf{0}, 2\eta^2\boldsymbol{S})$ and*

$\{L_n(\omega_j)\} \overset{A}{\sim} \{\frac{1}{2}\eta^2 S(\omega_j)Z_j\}$, *where* $\eta^2 := 1/(4f^2(0))$, $\boldsymbol{S} := \mathrm{diag}\{S(\omega_1), S(\omega_1), \ldots, S(\omega_q), S(\omega_q)\}$, *and* $Z_j \sim$ *i.i.d.* $\chi^2(2)$ $(j = 1, \ldots, q)$.

Proof. See Appendix. □

According to Theorem 3, the asymptotic mean of $L_n(\omega)$ is equal to $L(\omega) := \eta^2 S(\omega)$. We call this function the *Laplace spectrum* of $\{\varepsilon_t\}$. Recall that if $\{\varepsilon_t\}$ is stationary in second moments with autocorrelation function $\rho_\tau := \mathrm{Corr}(\varepsilon_{t+\tau}, \varepsilon_t)$, then the asymptotic mean of $G_n(\omega)$, known as the power spectrum (which may also be called the Gauss spectrum), is equal to $G(\omega) := \sigma^2 R(\omega)$, where $R(\omega) := \sum_{\tau=-\infty}^{\infty} \rho_\tau \cos(\omega\tau) = 2\sum_{\tau=0}^{\infty} \rho_\tau \cos(\omega\tau) - 1$ is the autocorrelation spectrum (or normalized power spectrum) of $\{\varepsilon_t\}$. Therefore, the Laplace spectrum $L(\omega)$ is the counterpart of the power spectrum $G(\omega)$; the former is proportional to the zero-crossing spectrum $S(\omega)$ as the latter is proportional to the autocorrelation spectrum $R(\omega)$. The zero-crossing spectrum is a Fourier transform of the zero-crossing rates, whereas the autocorrelation spectrum is a Fourier transform of the autocorrelation coefficients. As in the white noise case, η^2 is the counterpart of σ^2 as a multiplier.

The relationship between the autocorrelation coefficients and zero-crossing rates is complicated in general. But for some special distributions, it can be expressed in explicit forms. For example, if $\{\varepsilon_t\}$ is a stationary Gaussian process with autocorrelation function ρ_τ, then (Kedem 1994)

$$p_{ts} = \tfrac{1}{4}(1 - \delta_{t-s}) + \tfrac{1}{2}\pi^{-1} \arcsin(\rho_{t-s}) := p_{t-s}.$$

According to Lemma 1, $\{\varepsilon_t\}$ is also stationary in zero crossings with

$$\gamma_\tau = 1 - 2p_\tau = \tfrac{1}{2}(1 + \delta_\tau) - \pi^{-1} \arcsin(\rho_\tau).$$

Therefore, for stationary Gaussian processes, we have

$$S(\omega) = 1 + 4\pi^{-1} \sum_{\tau=1}^{\infty} \arcsin(\rho_\tau) \cos(\omega\tau), \tag{6}$$

which is a Fourier transform of arcsine-transformed autocorrelation coefficients.

The same expression can obtained for elliptical distribution. In fact, let $\boldsymbol{\varepsilon} := [\varepsilon_1, \ldots, \varepsilon_n]^T$ have an elliptical distribution with density $f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-1/2} g(\boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x})$ $(\boldsymbol{x} \in \mathbb{R}^n)$, where $\boldsymbol{\Sigma} := [\sigma_{ts}]_{t,s=1}^n$ is a positive-definite matrix with $\sigma_{tt} \equiv \sigma^2$ for all t and $g(u)$ is a nonnegative function satisfying $\int_{\boldsymbol{x} \in \mathbb{R}^n} g(\|\boldsymbol{x}\|^2)\, d\boldsymbol{x} = 1$. It can be shown that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \kappa\boldsymbol{\Sigma}$, where $\kappa := n^{-1} \int_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{x}\|^2 g(\|\boldsymbol{x}\|^2)\, d\boldsymbol{x}$. For any $u \geq 0$ and $1 \leq k < n$, let $g_k(u) := \int_{\boldsymbol{v} \in \mathbb{R}^{n-k}} g(u + \|\boldsymbol{v}\|^2)\, d\boldsymbol{v}$. Then, the marginal distribution of ε_t, for all t, can be expressed as $f(x) = \sigma^{-1/2} g_1(x^2/\sigma^2)$ and the joint distribution of ε_t and ε_s, for any $t \neq s$, takes the form $f_{ts}(\boldsymbol{x}) = |\boldsymbol{\Sigma}_{ts}|^{-1/2} g_2(\boldsymbol{x}^T \boldsymbol{\Sigma}_{ts}^{-1} \boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^2$, where

$$\boldsymbol{\Sigma}_{ts} := \begin{bmatrix} \sigma^2 & \sigma_{ts} \\ \sigma_{ts} & \sigma^2 \end{bmatrix}.$$

Since $(\varepsilon_t, \varepsilon_s)$ remains elliptically distributed, it can be shown (Kedem 1994) that

$$p_{ts} = \tfrac{1}{4}(1 - \delta_{t-s}) + \tfrac{1}{2}\pi^{-1} \arcsin(\rho_{ts}),$$

where $\rho_{ts} := \mathrm{Corr}(\varepsilon_t, \varepsilon_s) = \sigma_{ts}/\sigma^2$. Assume that $\{\varepsilon_t\}$ is stationary in second moments so that $\rho_{ts} = \rho_{t-s}$. Then, by Lemma 1, $\{\varepsilon_t\}$ is also stationary in zero crossings with $\gamma_\tau = 1 - 2p_\tau$. Consequently, we obtain the same expression (6) for the zero-crossing spectrum. Moreover, since $f(0) = \sigma^{-1/2} g_1(0)$, we have $\eta^2 = \sigma^2/(4g_1^2(0))$, where $\sigma^2 = V(\varepsilon_t)/\kappa$.

3.3 Time Series of Mixed Spectrum

Now consider the case where $\{y_t\}$ contains a sinusoidal signal $\mu_t = \boldsymbol{x}_t^T(\omega_0)\boldsymbol{\beta}_0$. A time series of this form is called a time series of mixed spectrum because its power spectrum is a mixture of discrete and continuous components. The following theorem establishes the asymptotic distribution of the Laplace periodogram at the signal frequency ω_0.

Theorem 4. *Let $\hat{\boldsymbol{\beta}}_n(\omega)$ be defined by (3) with $y_t = \boldsymbol{x}_t^T(\omega_0)\boldsymbol{\beta}_0 + \varepsilon_t$ $(t = 1,\ldots,n)$ for some constant $\omega_0 \in (0,\pi)$, where $\{\varepsilon_t\}$ satisfies the assumptions in Theorem 3. Then, as $n \to \infty$, $n^{1/2}(\hat{\boldsymbol{\beta}}_n(\omega_0) - \boldsymbol{\beta}_0) \overset{A}{\sim} N(\mathbf{0}, 2\eta^2 S(\omega_0)\mathbf{I})$ and $L_n(\omega_0) \overset{A}{\sim} \frac{1}{2}\eta^2 S(\omega_0)\chi^2(2, n\lambda)$, where $\eta^2 := 1/(4f^2(0))$ and $\lambda := \frac{1}{2}\|\boldsymbol{\beta}_0\|^2/(\eta^2 S(\omega_0))$.*

Proof. See Appendix. □

Recall that if $\{\varepsilon_t\}$ is stationary in second moments with autocorrelation spectrum $R(\omega)$, then the ordinary periodogram has the property that $G_n(\omega_0) \overset{A}{\sim} \frac{1}{2}\sigma^2 R(\omega_0)\chi^2(2, n\gamma)$, where $\gamma := \frac{1}{2}\|\boldsymbol{\beta}_0\|^2/(\sigma^2 R(\omega_0))$. Theorem 4 is the counterpart of this result for the Laplace periodogram. Since γ is known as the ordinary signal-to-noise ratio (SNR), we may refer to its counterpart $\lambda = \frac{1}{2}\|\boldsymbol{\beta}_0\|^2/(\eta^2 S(\omega_0))$ as the *Laplace SNR*. By Theorem 4, the Laplace SNR is a natural measure of the strength of the sinusoidal signal in the Laplace spectrum.

Next, let us consider the asymptotic distribution at non-signal frequencies. For simplicity, we focus on the case of white noise (a similar analysis can be carried out for colored noise).

Theorem 5. *Let $\hat{\boldsymbol{\beta}}_n(\omega)$ be defined by (3) with $y_t = \boldsymbol{x}_t^T(\omega_0)\boldsymbol{\beta}_0 + \varepsilon_t$ $(t = 1,\ldots,n)$, where $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with distribution function $F(x)$ and density $f(x)$. Let*

$\boldsymbol{W}_n$, $\boldsymbol{Q}_n$, and $\boldsymbol{h}_n$ be defined in Theorem 1 for $q = 1$ (with subscripts j and k omitted) with $w_t := -\boldsymbol{x}_t^T(\omega_0)\boldsymbol{\beta}_0$ and $\boldsymbol{x}_t := \boldsymbol{x}_t(\omega)$. Assume that there is a constant $\delta > 0$ such that

(viii) $f(x)$ is continuously differentiable for $|x| \le \|\boldsymbol{\beta}_0\| + \delta$ and $f(x) > 0$ for $|x| \le \|\boldsymbol{\beta}_0\|$.

Then, for $\omega \neq \omega_0$, $n^{1/2}\{\hat{\boldsymbol{\beta}}_n(\omega) - \boldsymbol{\beta}_n(\omega)\} \overset{A}{\sim} N(\boldsymbol{0}, 2\eta_1^2 \boldsymbol{V}_n(\omega))$, where $\boldsymbol{\beta}_n(\omega) := n^{-1/2}\boldsymbol{Q}_n^{-1}\boldsymbol{h}_n$, $\boldsymbol{V}_n(\omega) := (2\eta_1^2)^{-1}\boldsymbol{Q}_n^{-1}\boldsymbol{W}_n\boldsymbol{Q}_n^{-1}$, and $\eta_1^2 := c_0/b_0^2$, with $b_0 := n^{-1}\sum_{t=1}^n f(w_t)$ and $c_0 := n^{-1}$ $\sum_{t=1}^n F(w_t)\{1 - F(w_t)\}$. Moreover, let $\boldsymbol{V}_n(\omega) = \boldsymbol{U}_n\boldsymbol{\Sigma}_n\boldsymbol{U}_n^T$ be the singular-value decomposition, where $\boldsymbol{U}_n$ is an unitary matrix and $\boldsymbol{\Sigma}_n := \mathrm{diag}\{S_1(\omega), S_2(\omega)\}$. Then, for $\omega \neq \omega_0$, $L_n(\omega) \overset{A}{\sim} \frac{1}{2}\eta_1^2\{S_1(\omega)Z_1^2 + S_2(\omega)Z_2\}$, where $Z_1 \sim \chi^2(1, \lambda_1(\omega))$ and $Z_2 \sim \chi^2(1, \lambda_2(\omega))$ are independent, $\lambda_i(\omega) := \mu_i^2(\omega)/(2\eta_1^2 S_i(\omega))$ $(i = 1, 2)$, and $[\mu_1(\omega), \mu_2(\omega)]^T := \boldsymbol{U}_n^T\boldsymbol{\beta}_n(\omega)$.

Proof. See Appendix □

To appreciate the implications of Theorem 5, let us evaluate the quantities involved under the additional assumption that $\omega_0 = 2\pi k_0/n$ and $\omega = 2\pi k/n$ for some integers k_0 and k such that $0 < k_0 \neq k < \frac{1}{2}n$ and $n/k_0 = 2m+1$ for some integer $m > 0$. In this case, it can be shown (see Appendix) that

$$\boldsymbol{\beta}_n(\omega) = b_0^{-1}\boldsymbol{B}^{-1}\boldsymbol{a}, \quad \boldsymbol{V}_n(\omega) = \boldsymbol{B}^{-1}\boldsymbol{C}\boldsymbol{B}^{-1}, \tag{7}$$

where $\boldsymbol{a} := \sum_{\ell=1}^m \delta_{\ell k_0 - k}\boldsymbol{a}_\ell$, $\boldsymbol{B} := \mathbf{I} + \sum_{\ell=1}^m (\delta_{\ell k_0 - 2k} + \delta_{n-\ell k_0 - 2k})\boldsymbol{B}_\ell$, and $\boldsymbol{C} := \mathbf{I} + \sum_{\ell=1}^m (\delta_{\ell k_0 - 2k} + \delta_{n-\ell k_0 - 2k})\boldsymbol{C}_\ell$ for some $\boldsymbol{a}_\ell$, $\boldsymbol{B}_\ell$, and $\boldsymbol{C}_\ell$. This result reveals a possibility of leak from line spectrum in the Laplace periodogram. The spectral leak manifests itself in $\boldsymbol{\beta}_n(\omega)$ at the harmonic frequencies of the sinusoidal signal and in $\boldsymbol{V}_n(\omega)$ at the semi-harmonic frequencies.

More specifically, we need to distinguish three different cases. Case 1, $\omega = 2\pi k/n$ is not a harmonic or semi-harmonic frequency, i.e., $k \notin \mathcal{H} := \{\frac{1}{2}\ell k_0 \colon \ell = 1, \ldots, 2m; \ell \neq 2\}$; we call

it a type I frequency. In this case, $\boldsymbol{a} = \mathbf{0}$, $\boldsymbol{B} = \boldsymbol{C} = \mathbf{I}$, so $\boldsymbol{\beta}_n(\omega) = \mathbf{0}$ and $\boldsymbol{V}_n(\omega) = \mathbf{I}$. Therefore, we have $n^{1/2}\hat{\boldsymbol{\beta}}_n(\omega) \overset{A}{\sim} N(\mathbf{0}, 2\eta_1^2\mathbf{I})$ and

$$L_n(\omega) \overset{A}{\sim} \tfrac{1}{2}\eta_1^2\chi^2(2). \tag{8}$$

This means that at type I frequencies the Laplace periodogram has the same asymptotic distribution as in the signal-free case except that the scaling factor is $\frac{1}{2}\eta_1^2$ instead of $\frac{1}{2}\eta^2$. The spectral leak manifests itself in η_1^2 as a signal-dependent quantity.

Case 2, ω is a semi-harmonic frequency, i.e., $k = \frac{1}{2}\ell k_0$ for some odd integer $\ell \in [1, 2m]$; we call it a type II frequency. This is possible only if k_0 is an even integer. In this case, $\boldsymbol{a} = \mathbf{0}$, $\boldsymbol{B} = \mathbf{I} + \boldsymbol{B}_\ell$, and $\boldsymbol{C} = \mathbf{I} + \boldsymbol{C}_\ell$, so that $n^{1/2}\hat{\boldsymbol{\beta}}_n(\omega) \overset{A}{\sim} N(\mathbf{0}, 2\eta_1^2\boldsymbol{V}_n(\omega))$, where $\boldsymbol{V}_n(\omega) = (\mathbf{I} + \boldsymbol{B}_\ell)^{-1}(\mathbf{I} + \boldsymbol{C}_\ell)(\mathbf{I} + \boldsymbol{B}_\ell)^{-1}$. Because $\boldsymbol{V}_n(\omega)$ is not necessarily proportional to $\mathbf{I}$, the Laplace periodogram at type II frequencies does not generally have a scaled central $\chi^2(2)$ distribution. Instead, it is distributed as a weighted sum of two independent central $\chi^2(1)$ random variables. Even in the special case where $b_{2\ell} = c_{2\ell} = 0$, we only have $\boldsymbol{V}_n(\omega) = \operatorname{diag}(v_{11}, v_{22})$, where $v_{11} := (1 + \frac{1}{2}c_{1\ell}/c_0)/(1 + \frac{1}{2}b_{1\ell}/b_0)^2$ and $v_{22} := (1 - \frac{1}{2}c_{1\ell}/c_0)/(1 - \frac{1}{2}b_{1\ell}/b_0)^2$ are not necessarily equal. Nonetheless, we can assume $\boldsymbol{V}_n(\omega) = v\mathbf{I}$ to obtain a first-order approximation

$$L_n(\omega) \overset{A}{\sim} \tfrac{1}{2}\eta_1^2 v\chi^2(2). \tag{9}$$

The extra factor v distinguishes (9) from (8). By setting $v := \frac{1}{2}(v_{11} + v_{22})$, the distribution in (9) matches the exact asymptotic mean of the Laplace periodogram, which equals $\eta_1^2 v$. The signal dependence of the scaling factor is a manifestation of spectral leak from the signal.

Case 3, ω is a harmonic frequency, i.e., $k = \ell k_0$ for some integer $\ell \in [2, m]$; we call it a type III frequency. At such a frequency, $\boldsymbol{\beta}_n(\omega)$ is nonzero and (7) cannot be further simplified,

so the Laplace periodogram is distributed as a weighted sum of two independent noncentral $\chi^2(1)$ random variables. As a first-order approximation, we can write

$$L_n(\omega) \overset{A}{\sim} \tfrac{1}{2}\eta_1^2 v \chi^2(2, n\lambda_1), \tag{10}$$

where $\lambda_1 := \frac{1}{2}\|\boldsymbol{\beta}_n(\omega)\|^2/(\eta_1^2 v)$. Recall that the Laplace periodogram at the signal frequency ω_0 has the following distribution:

$$L_n(\omega_0) \overset{A}{\sim} \tfrac{1}{2}\eta^2 \chi^2(2, n\lambda), \tag{11}$$

where $\eta^2 = 1/(4f^2(0))$ and $\lambda = \frac{1}{2}\|\boldsymbol{\beta}_0\|^2/\eta^2$. The distribution in (10) is similar to (11) except that the scaling factor and the noncentrality parameter are $\frac{1}{2}\eta_1^2 v$ and λ_1 instead of $\frac{1}{2}\eta^2$ and λ. The mean of this distribution coincides with the exact asymptotic mean of $L_n(\omega)$, which equals $\frac{1}{4}n\|\boldsymbol{\beta}(\omega)\|^2 + \eta_1^2 v$. Note that the presence of a nonzero noncentrality parameter implies that a peak may appear at type III frequencies, so interpretation of spectral peaks must be done with caution.

4. SIMULATION STUDIES

In this section, we provide some simulation examples of practical importance to demonstrate the benefit of the Laplace periodogram over the ordinary periodogram and to validate the theoretical findings discussed in the previous section.

4.1 Robust Spectral Estimation

The first example is a validation of the Laplace periodogram as a viable alternative to the ordinary periodogram for analyzing time series of continuous spectrum.

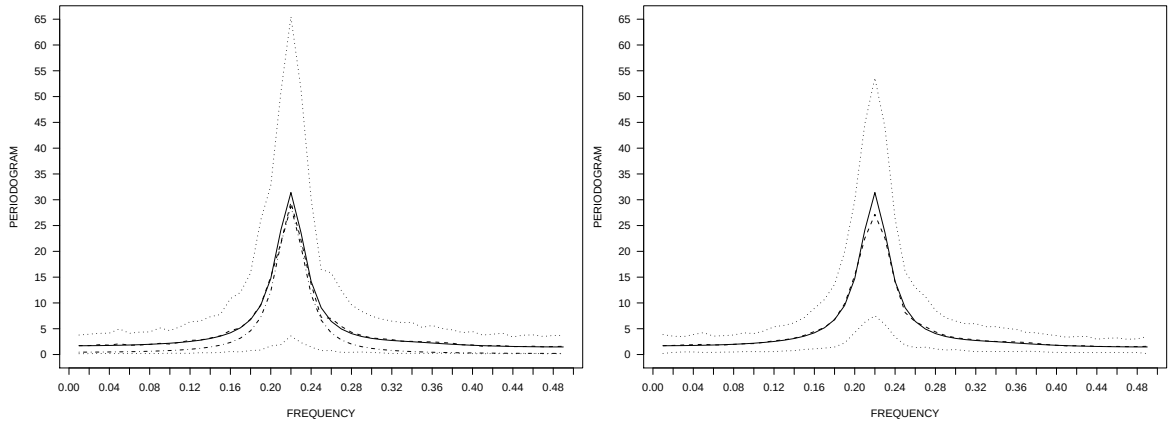
Example 1. Let $y_t = \varepsilon_t$ $(t = 1, \ldots, n)$, where $\{\varepsilon_t\}$ is a Gaussian AR(2) process satisfying $\varepsilon_t = a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2} + \xi_t$, $a_1 := 2r\cos(\omega_c)$, $a_2 := -r^2$, $\{\xi_t\} \sim$ i.i.d. $N(0,1)$. It can be shown that $\sigma^2 := V(\varepsilon_t) = (1-a_2)\{(1+a_2)((1-a_2)^2 - a_1^2)\}^{-1}$ and

$$\rho_\tau = \frac{\sin((\tau+1)\omega_c) - r^2\sin((\tau-1)\omega_c)}{(1+r^2)\sin(\omega_c)}\, r^\tau \quad (\tau \geq 0).$$

Figure 1 shows a simulation result on the basis of 500 Monte Carlo runs with $r = 0.9$, $\omega_c = 0.22 \times 2\pi$, and $n = 100$. In this example, the Laplace spectrum, calculated according to (6), has the same bandpass characteristics as the power spectrum, and the simulated mean of the Laplace periodogram follows the Laplace spectrum closely. Figure 1 also shows that smoothing does help reduce the statistical variability of the Laplace periodogram as an estimator of the Laplace spectrum.

According to Theorem 3, one of the advantages of the Laplace spectrum over the ordinary power spectrum is that the Laplace spectrum is invariant (up to a constant multiplier) to any memoryless nonlinearity that preserves the signs of the time series data. A practical example of such nonlinearities is the nonlinear and clipping distortions in data acquisition and transmission systems (Bahai et al. 2002; Chorti and Brookes 2006). This type of nonlinearity is known to cause distortions in the power spectrum (Wise, Traganitis, and Thomas 1977) and cannot be modeled as additive noise. The Laplace periodogram is an effective tool for handling such nonlinearity, as demonstrated in the next example.

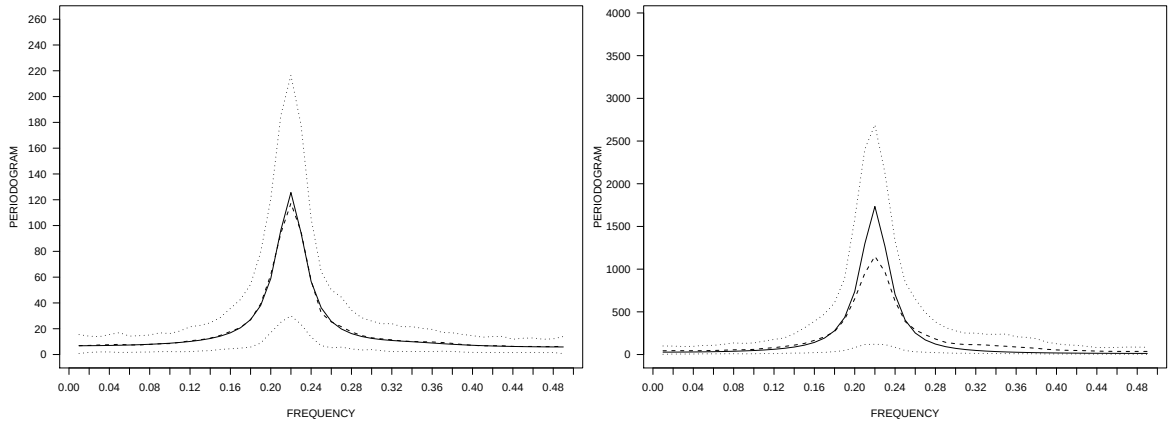
Example 2. Let $y_t = \varphi(\varepsilon_t)$ $(t = 1, \ldots, n)$, where $\{\varepsilon_t\}$ is the process discussed in Example 1 and $\varphi(x)$ is a sign-preserving function satisfying $\varphi(0) = 0$ and $\dot{\varphi}(0) \neq 0$. Because the transform does not alter the zero-crossing spectrum, nor does it violate the assump-



(a) (b)

Figure 1: Laplace periodogram as function of $\omega/(2\pi)$ for the AR(2) process in Example 1. Solid line, true Laplace spectrum; dashed line, sample mean of 500 independent Laplace periodograms; dotted line, 10th and 90th pointwise percentiles of Laplace periodograms; dash-dotted line in (a), true power spectrum. (a) Raw periodograms (RMSE=0.945). (b) Smoothed periodograms by smoothing splines (RMSE=0.676).

tions in Theorem 3, the Laplace periodogram of $\{\varphi(\varepsilon_t)\}$ has the same asymptotic distribution as $\{\varepsilon_t\}$, except that $f(0)$ should be replaced by $f(0)/\dot{\varphi}(0)$ in calculating η^2 so that $\eta^2 = \dot{\varphi}^2(0)/(4f^2(0))$. Figure 2 shows a simulation result with $\varphi(x) := 2xI(|x| \leq \sigma) + 2(x^3/\sigma^2)I(|x| > \sigma)$, where $\sigma^2 := V(\varepsilon_t)$. In this case, the Laplace spectrum is simply a constant multiple of the Laplace spectrum shown in Figure 1, with multiplier $\dot{\varphi}^2(0) = 4$. This spectrum, depicted in Figure 2(a) by the solid line, is again followed closely by the simulated mean of the Laplace periodogram across all frequencies. In contract, the ordinary periodogram is considerably distorted by the nonlinearity, as shown in Figure 2(b), resulting in a lowered spectral peak. Note that the power spectrum in Figure 2(b) (solid line) is scaled by the variance of $\{y_t\}$ instead of σ^2.

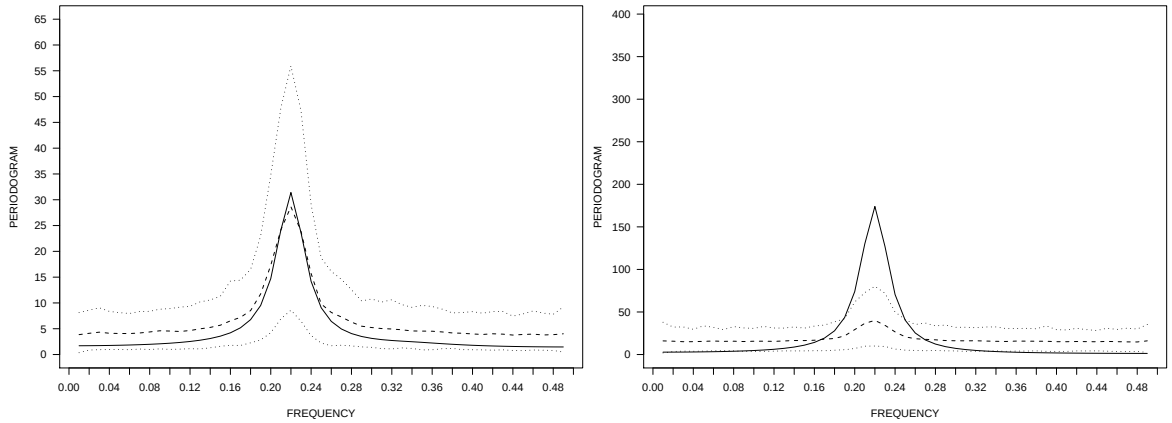


(a) (b)

Figure 2: (a) Laplace periodogram (RMSE=0.677), similar to Figure 1(b) except that the data are transformed by the memoryless nonlinearity discussed in Example 2. (b) Ordinary periodogram for the same data (RMSE=2.01), with solid line representing the (scaled) power spectrum of the untransformed process.

Because of its origin in the LAD regression, the Laplace periodogram is expected to be more robust against impulsive contaminations (outliers) in the data than the ordinary periodogram. The next example confirms this advantage of the Laplace periodogram.

Example 3. Consider again the AR(2) process in Example 1 and suppose that the observations of the process are contaminated by impulsive noise. In particular, assume that a randomly selected $100p\%$ of the data points are contaminated by additive i.i.d. $N(0,\sigma_c^2)$ noise with some $\sigma_c^2 \gg \sigma^2$. Figure 3 shows a simulation result with $p = 0.2$ and $\sigma_c = 5\sigma$. As can be seen, the impulsive noise has a mild impact on the Laplace periodogram but a dramatic impact on the ordinary periodogram: the spectral peak remains prominent in the Laplace periodogram, but almost disappears in the ordinary periodogram.



(a) (b)

Figure 3: (a) Laplace periodogram (RMSE=1.01), similar to Figure 1(b) except that the data are contaminated by the impulsive noise discussed in Example 3 and the solid line represents the Laplace spectrum of the uncontaminated process. (b) Ordinary periodogram for the same data (RMSE=6.28), similar to Figure 2(b) except that the solid line represents the (scaled) power spectrum of the uncontaminated process.

Since the Laplace spectrum is proportional to the zero-crossing spectrum and the latter, according to Lemma 1, can be estimated directly by 4 times the ordinary periodogram of the binary time series $I(y_t < 0)$ $(t = 1, \ldots, n)$, we can form a simple estimator of the Laplace spectrum by multiplying the ordinary periodogram of $\{I(y_t < 0)\}$ with an estimator of $4\eta^2$. This simple method has the computational advantage over the Laplace periodogram, because the ordinary periodogram can be easily computed by a fast Fourier transform. However, the simple method may suffer from a loss of statistical efficiency.

To compare these methods in terms of statistical efficiency, we resort to a simulation

Table 1: Laplace Periodogram and Simple Estimator

	Laplace Periodogram		*Simple Estimator*	
Case	RMSE-1	RMSE-2	RMSE-1	RMSE-2
$r = 0$	1.56	0.99	1.57	1.00
$r = 0.50$	1.86	1.18	1.91	1.22
$r = 0.90$	11.50	4.45	12.40	4.92
$r = 0.95$	31.17	8.15	34.36	9.26

Results are based on 5,000 Monte Carlo runs.

study in which two types of mean-squared error,

$$\text{MSE-1} := m^{-1} \sum_{k=1}^{m} E\{|\hat{L}_n(\omega_k) - L(\omega_k)|^2\} \quad \text{and}$$
$$\text{MSE-2} := m^{-1} \sum_{k=1}^{m} E\{|\hat{L}_n(\omega_k)/L(\omega_k) - 1|^2\}$$

with $\omega_k := 2\pi k/n$ and $m := \lfloor \frac{1}{2} n \rfloor$, are computed from Monte Carlo samples. Table 1 contains the root mean-squared error (RMSE) from the simulation study for the AR(2) process defined in Example 1 with $\omega_c = 0.22 \times 2\pi$ and $n = 100$. In the simple estimator, η^2 is estimated by 0.5π times the sample variance (because $\eta^2 = \frac{1}{2}\pi\sigma^2$ for Gaussian processes). As can be seen, the Laplace periodogram outperforms the simple estimator significantly in all cases except for the case of white noise ($r = 0$) where the two estimators perform similarly. The margin in favor of the Laplace periodogram becomes more pronounced as the spectral peak sharpens (when r approaches unity).

Based on the simulation studies in this section, we conclude that the Laplace periodogram offers a robust but effective alternative to the ordinary periodogram for analyzing time series of continuous spectrum. The robustness comes in two forms: robustness against impulsive

noise (outliers) and robustness against memoryless nonlinearity. Both have important practical benefits in applications.

4.2 Signal Detection

Consider the problem of detecting a sinusoidal signal $\mu_t = \boldsymbol{x}_t^T(\omega_0)\boldsymbol{\beta}_0$ from noisy observations $y_t = \mu_t + \varepsilon_t$ $(t = 1, \ldots, n)$. We assume that the frequency ω_0 is known but the amplitude $\boldsymbol{\beta}_0$ is not. This problem can be formulated as that of testing the hypotheses $H_0 : \boldsymbol{\beta}_0 = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_0 \neq \mathbf{0}$. We do not assume any knowledge of the noise $\{\varepsilon_t\}$ except for η^2 and $S(\omega_0)$. These parameters can be estimated in practical situations where observations can be made in the absence of the signal.

Under these conditions, a natural detector, which we call the Laplace detector, is

$$L_n(\omega_0) \begin{cases} > \theta & H_1\text{: signal present,} \\ \leq \theta & H_0\text{: signal absent.} \end{cases}$$

This detector is similar to an F test in spirit, except that the regression coefficients are not standardized by an estimate of the noise variance. Under H_0, we have $L_n(\omega_0) \overset{A}{\sim} \frac{1}{2}\eta^2 S(\omega_0)$. Therefore, the false-alarm probability of the Laplace detector can be made approximately equal to $\alpha \in (0,1)$ by setting the threshold $\theta := \frac{1}{2}\eta^2 S(\omega_0)\chi^2_{1-\alpha}(2)$, where $\chi^2_{1-\alpha}(2)$ denotes the $(1-\alpha)$ quantile of the central $\chi^2(2)$ distribution. According to Theorem 4, the detection probability of the Laplace detector under H_1 is approximately $P_L(\alpha) := P\{\chi^2(2, n\lambda) > \chi^2_{1-\alpha}(2)\}$. A similar detector, which we call the Gauss detector, can be derived from the ordinary periodogram. The detection probability of the Gauss detector is approximately $P_G(\alpha) := P\{\chi^2(2, n\gamma) > \chi^2_{1-\alpha}(2)\}$.

To compare the power of these detectors, we observe that $P\{\chi^2(2,u) > \chi^2_{1-\alpha}(2)\}$ is an

increasing function of $u \geq 0$. This implies that $P_L(\alpha) > P_G(\alpha) \Leftrightarrow \lambda > \gamma$. In other words, the Laplace detector is more powerful than the Gauss detector if and only if the Laplace SNR is greater than the ordinary SNR. By definition, $\lambda > \gamma$ is equivalent to $\eta^2 S(\omega_0) < \sigma^2 R(\omega_0)$, and the latter depends solely on the noise spectrum at the signal frequency. In the special case of white noise, this condition reduces to $\eta^2 < \sigma^2$. For Gaussian white noise, $\eta^2 = \frac{1}{2}\pi\sigma^2 > \sigma^2$, so the Gauss detector is more powerful in this case. If the noise has a double exponential distribution, then $\eta^2 = \frac{1}{2}\sigma^2 < \sigma^2$, in which case the Laplace detector becomes more powerful. Generally, because $\eta^2 < \sigma^2$ is equivalent to $\sigma f(0) > \frac{1}{2}$, this condition favors heavy-tailed distributions with extremely large σ relative to the density at zero.

As a numerical example, we present a simulation result in Figure 4, where the receiver operating characteristic (ROC) curves of the Laplace and Gauss detectors are obtained from 1,000 Monte Carlo runs with $n = 100$ under different noise conditions (both false alarm and detection probabilities are simulated). In (a)(b), $\boldsymbol{\beta}_0 = [0.5, 0]^T$, $\omega_0 = 0.15 \times 2\pi$, and $\{\varepsilon_t\}$ is an i.i.d. sequence (white noise). In (c)(d), $\boldsymbol{\beta}_0 = [1, 0]^T$, $\omega_0 = 0.15 \times 2\pi$, and $\{\varepsilon_t\}$ is an AR(2) process (colored noise) defined in Example 1 with $r = 0.8$, $\omega_c = 0.18 \times 2\pi$, and driven by different white noise. The variance of the white noise equals unity in all but the Cauchy case. For the Cauchy white noise, the scale parameter equals 0.25.

As can be seen from Figure 4, the Gauss detector outperforms the Laplace detector in the Gaussian case. This is expected because of the maximum likelihood interpretation of the Gauss detector. But in all other cases, where the noise has heavier tails, the Laplace detector is superior. Particularly in the Cauchy case, the noise does not have a finite variance, but the Laplace detector performs remarkably well against the Gauss detector. This result confirms

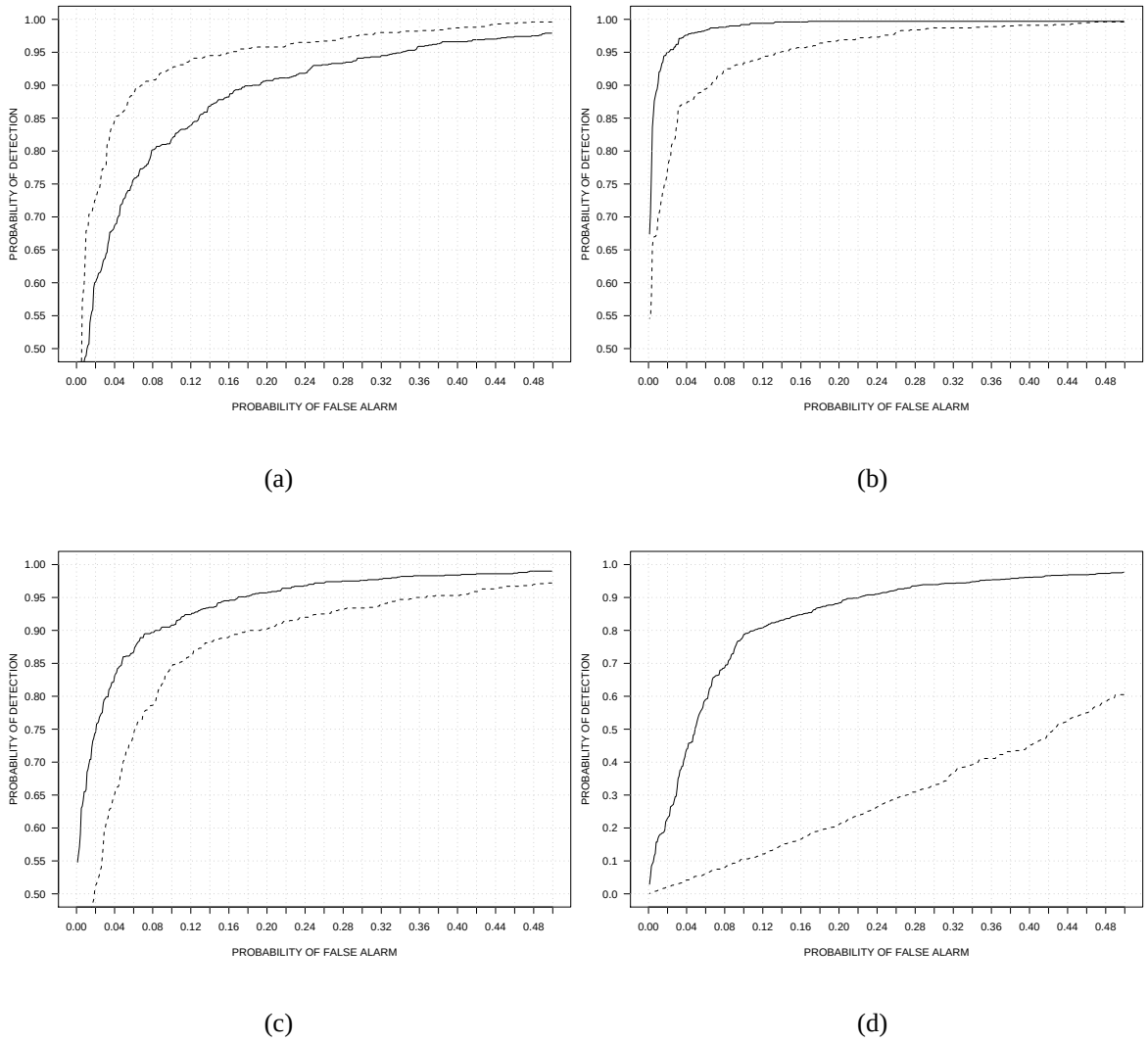


Figure 4: ROC curves for detecting a sinusoidal signal under different noise conditions. (a) Gaussian white noise; (b) double exponential white noise; (c) AR(2) driven by Student's *T* white noise with 3 degrees of freedom; (d) AR(2) driven by Cauchy white noise. Solid line, Laplace detector; dashed line, Gauss detector.

Table 2: $\boldsymbol{\beta}_n(\omega)$ and $\boldsymbol{V}_n(\omega)$ at Type II and III Frequencies

	$k=5$	$k=15$	$k=20$	$k=25$	$k=30$
β_1	0.00000	0.00000	0.04555	0.00000	−0.37487
v_{11}	1.00331	0.94119	0.94119	1.57140	1.00331
v_{22}	0.99643	1.06166	1.06166	0.72580	0.99643

the superiority of the Laplace periodogram for signal detection in heavy-tailed noise.

4.3 Mixed Spectrum

Consider the time series of mixed spectrum $y_t = \boldsymbol{x}_t^T(\omega_0)\boldsymbol{\beta}_0 + \varepsilon_t$ $(t = 1, \ldots, n)$, where $\{\varepsilon_t\}$ is white noise. To study the spectral leak discussed in Section 3.3, Figure 5 depicts $\boldsymbol{\beta}_n(\omega)$ and $\boldsymbol{V}_n(\omega)$ as functions of k (with $\omega = 2\pi k/n$) under the assumption of Gaussian white noise with unit variance and for $\omega_0 = 2\pi k_0/n$, $\boldsymbol{\beta}_0 = [2,0]^T$, $k_0 = 10$, and $n = 70$. In this example, type II frequencies correspond to $k = 5, 15, 25$ and type III frequencies correspond to $k = 20, 30$; all remaining non-signal frequencies are type I frequencies. Table 2 contains the numerical values of the first element of $\boldsymbol{\beta}_n(\omega)$ and the diagonal elements of $\boldsymbol{V}_n(\omega)$ at type II and III frequencies. We can see from Figure 5 that $\boldsymbol{\beta}_n(\omega) = \mathbf{0}$ and $\boldsymbol{V}_n(\omega) = \mathbf{I}$ at all type I frequencies. We can see from Table 2 that $\boldsymbol{V}_n(\omega)$ is not proportional to $\mathbf{I}$ at all type II and III frequencies and that $\boldsymbol{\beta}_n(\omega) \neq \mathbf{0}$ at all type III frequencies.

To validate the asymptotic (analytical) distributions in (8)–(11), we resort to a simulation in which simulated distributions of the Laplace periodogram at different types of frequencies and under various noise models. Figure 6 shows some of the results for Gaussian white noise with unit variance based on 1,000 Monte Carlo runs. As can be seen, the exact asymptotic

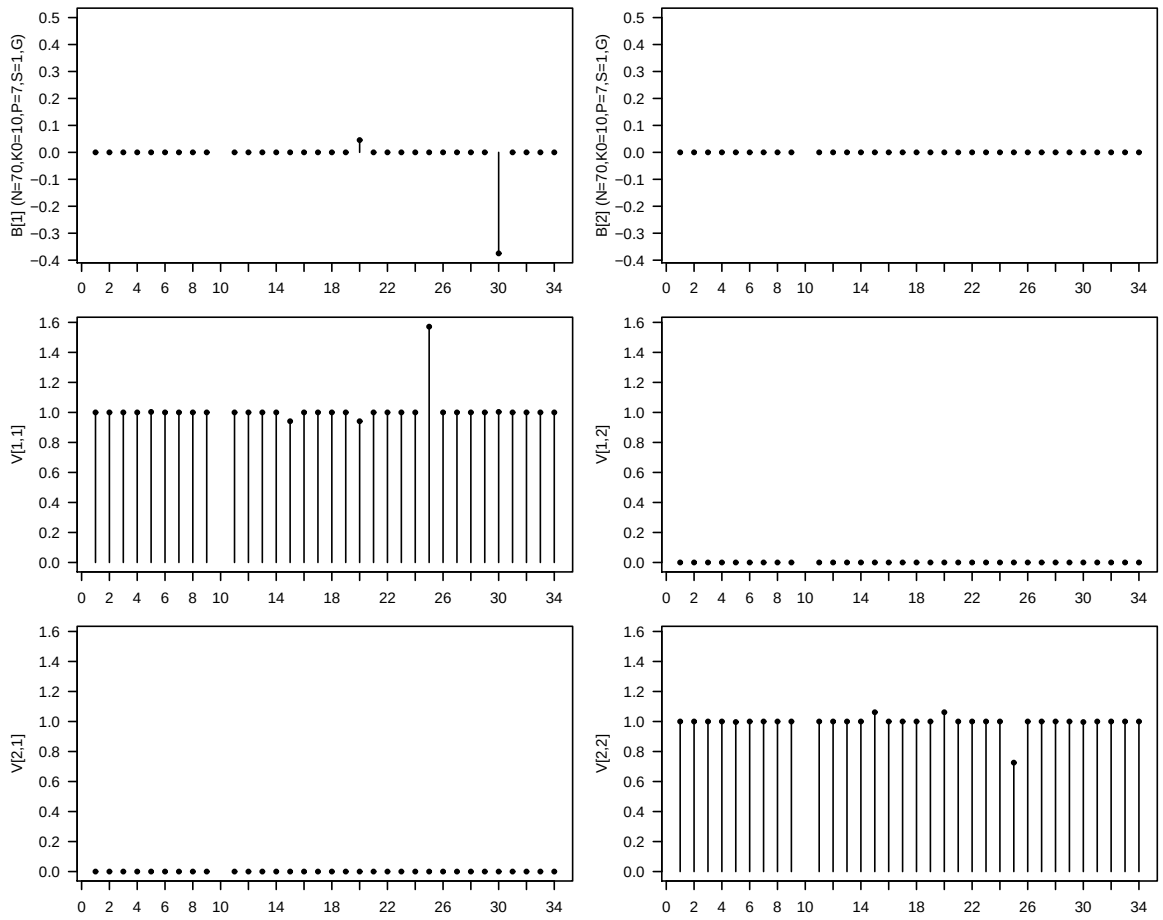


Figure 5: Plot of $\boldsymbol{\beta}_n(\omega)$ and $\boldsymbol{V}_n(\omega)$ in (7) at $\omega = 2\pi k/n$ ($k = 1, \ldots, \frac{1}{2}n - 1; k \neq k_0$) for a time series of mixed spectrum ($n = 70$ and $k_0 = 10$).

distributions in (8) and (11) are reasonably accurate even for moderate samples sizes such as $n = 70$; the approximate distributions in (9) and (10) do not work as well sometimes, especially when the magnitude of $\boldsymbol{\beta}_n(\omega)$ is large (e.g., $k = 25$) and/or the disparity between v_{11} and v_{22} is large (e.g., $k = 30$).

Under the same model, a comparison is made between the Laplace periodogram and the ordinary periodogram. Figure 7(a) depicts the simulated mean and 10th-90th percentiles of the Laplace periodogram based on 1,000 Monte Carlo runs. Figure 7(b) shows the corre-

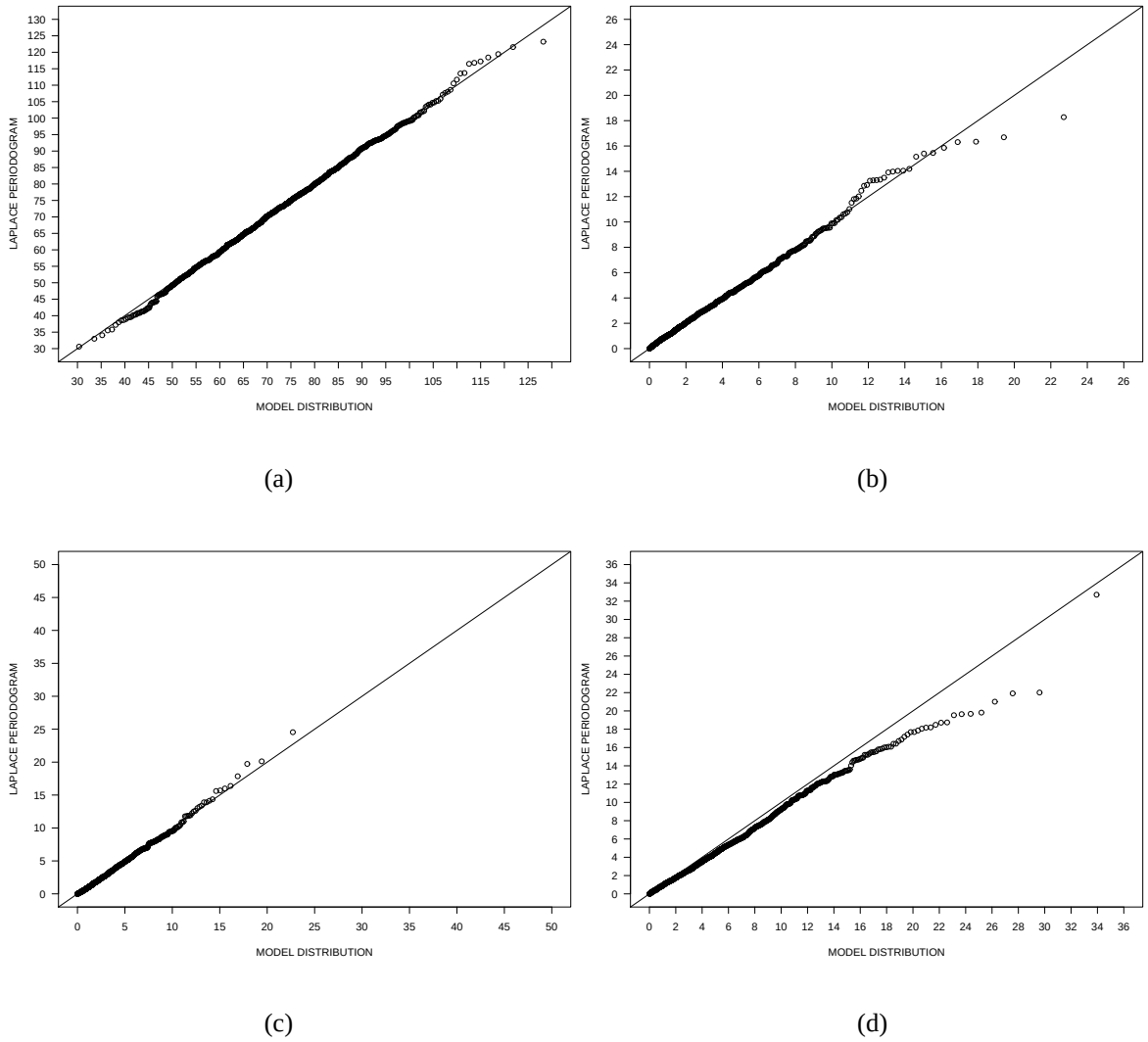


Figure 6: Q-Q plot of simulated versus analytical distributions of the Laplace periodogram at selected frequencies for a time series of mixed spectrum. (a) distribution (11) at signal frequency ($k = 10$); (b) distribution (8) at type I frequency ($k = 9$); (c) distribution (9) at type II frequency ($k = 5$); (b) distribution (10) at type III frequency ($k = 30$).

sponding results for the ordinary periodogram. As can be seen, both periodograms consist of a strong peak at the signal frequency ($k = 10$) and a constant noise floor. Due to the spectral leak, the Laplace periodogram also has a small but visible peak at $k = 30$.

Although the spectral leak puts the Laplace periodogram at a disadvantage, a real benefit of the Laplace periodogram shines through when comparing Figure 7(c) and (d), where the noise has a Cauchy distribution with unit scale parameter. In this case, the Laplace periodogram retains its characteristics of the Gaussian case, but the ordinary periodogram collapses miserably: the simulated mean and 90th percentile are extremely large at all frequencies with no visible clue about the location of the sinusoidal signal.

From this experiment, we conclude that the robustness benefit of the Laplace periodogram can justify its usefulness for time series of mixed spectrum provided that the possibility of spectral leak is well understood.

4.4 Frequency Estimation

Finally, we consider the classical problem of estimating the frequency of a sinusoidal signal from noisy observations $y_t = A\cos(\omega_0 t) + B\sin(\omega_0 t) + \varepsilon_t$ $(t = 1, \ldots, n)$, where A, B, and ω_0 are unknown constants and $\{\varepsilon_t\}$ is a zero-mean random process of unknow strength. Traditionally, the frequency is estimated by maximization of the ordinary periodogram (Quinn and Hannan 2001). But in situations where the noise has a heavy-tailed distribution, the Laplace periodogram provides an effective alternative to the traditional approach.

Table 3 contains the result of a simulation study where the RMSE for estimating the unknown frequency $\omega_0/(2\pi)$ is calculated on the basis of 1,000 Monte Carlo runs for the

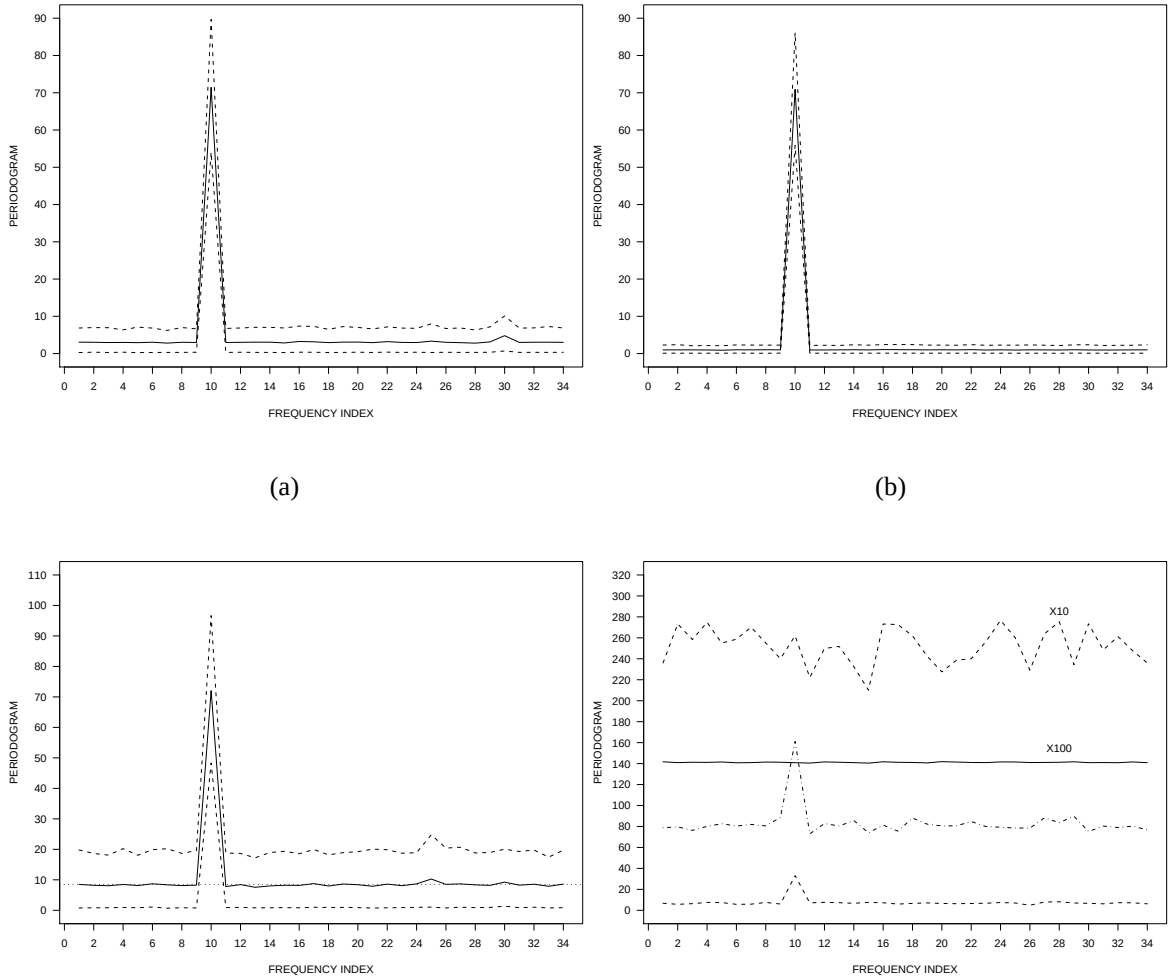


Figure 7: (a)(c) Laplace periodogram for a time series of mixed spectrum. (b)(d) Ordinary periodogram for the same data. Solid line, mean; dashed line, 10th and 90th percentile; dash-dotted line in (d), median. The noise is Gaussian in (a)(b) and Cauchy in (c)(d).

Table 3: RMSE of Frequency Estimation

	$n = 50$	$n = 100$	$n = 20$	$n = 300$	$n = 500$	$n = 900$
Laplace Periodogram	8.76e-03	2.22e-03	3.57e-04	1.45e-04	8.53e-05	5.69e-05
Ordinary Periodogram	9.83e-03	3.66e-03	6.52e-04	3.13e-04	1.85e-04	1.25e-04

maximizer of the Laplace periodogram and that of the ordinary periodogram. In this simulation, $A = 1$, $B = 0$, and $\omega_0 = 0.152 \times 2\pi$. The noise is a normalized Cauchy i.i.d. sequence, normalized for each realization so that the sample variance of the noise equals 0.5 and hence the signal-to-noise ratio equals 1 (0 dB). Clearly, the maximizer of the Laplace periodogram is a more accurate estimator in this case for all the sample sizes.

5. ANALYSIS OF HEART RATE VARIABILITY

Heart rate variability (HRV) is a measure of variations in heart rate over time from the QRS complexes in a continuous electrocardiographic (ECG) record. It is typically represented as a time series of beat-to-beat (RR) intervals where "time" refers to the cardiac cycle or heartbeat number. Spectral analysis techniques have been used to study the relationship between HRV and various physiological conditions because HRV reflects the activity of the autonomic nervous system (ANS) in regulating the sinus rhythm (Malik and Camm 1995). Here, we are not concerned with physiological explanations of HRV. Our focus is on the signal processing aspect of HRV analysis. In particular, we are interested in dealing with the inevitable problem of contamination by ectopic heartbeats (also known as premature ventricular contractions, or PVCs) and some artifacts capable of bankrupting conventional methods

(Albrecht and Cohen 1988; Malik et al. 1996).

Because of premature or skipped heartbeats, ectopic events are often manifested as sharp spikes in the RR tachogram (a plot of RR intervals against the beat number). Failure in beat detection during data aquisition also introduces spiky artifacts (missing beats). The ordinary power spectrum is ill-suited for handling such contaminations because of their significant power contribution. In practice, if these abnormal events are very rare, as in short-term studies, one simply removes them and interpolates the vacancies by adding phantom beats where sinus beats would have been expected to occur (Lippman, Stein, and Lerman 1994). If the occurrence of abnormal events is high, the entire segments that contain the events have to be eliminated from the analysis. This means that in many cases a great deal of data have to be discarded, resulting in severe restrictions on long-term study.

Although attempt has been made to design algorithms that automatically reject artifacts and interpolate ectopic beats (Acar et al. 2000; Storck et al. 2001), these algorithms are not error-free or harmless (Clifford and Tarassenko 2005). So it is highly desirable to have robust analysis methods that are insensitive to the spikes so that HRV can be analyzed effectively in the presence of ectopic beats and artifacts (Mateo and Laguna 2003). The Laplace periodogram is a promising candidate for this task, especially in light of its great robustness properties demonstrated in the previous section. In this section, we apply the Laplace periodogram to a long-term HRV series that contains many ectopic beats and artifacts.

The data are obtained from the CAST RR Interval Sub-Study Database (file m003b) hosted in PhysioNet (www.physionet.org/physiobank/database/crisdb), a web site that offers free access to large collections of recorded physiological signals. Of course, our purpose is

not to study the effectiveness of antiarrhythmic drugs, which is originally the intended use of the database. Our purpose is to demonstrate the effectiveness of the Laplace periodogram in analyzing HRV signals in the presence of ectopic events and artifacts.

Figure 8 shows the RR tachogram and its two spectrograms: the Laplace spectrogram and the ordinary spectrogram. The spectrograms depict the spectral evolution of the RR intervals over time. They are created from the corresponding periodograms calculated locally using the data points (multiplied by Hamming window after median removal) that fall into a moving window of length $n = 256$ beats (roughly 4 minutes) with an overlap of 32 beats (about half a minute). Large values in the spectrograms are coded by bright colors.

As can be seen, the ordinary spectrogram is completely masked whenever a spike falls into the moving window. Due to the frequent occurrence of such contaminative events, the ordinary spectrogram becomes useless most of the time. In contrast, the Laplace spectrogram remains largely intact during these events. As a result, the narrowband low-frequency (LF) component around 0.06 remains visible almost all the time. This component coincides with a similar component in the ordinary spectrogram of the spike-free segments. Moreover, the Laplace spectrogram also reveals some broadband high-frequency (HF) activities in the frequency range of 0.2–0.4. The bright vertically lines that run across the frequency span with a strong presence in the very-low-frequency (VLF) range are well correlated with persistent increasing or decreasing transitions in the heart rate (i.e., local trends in the RR tachogram).

The destructive impact of spiky contamination on the ordinary periodogram and the robustness of the Laplace periodogram to such contamination can be further appreciated by comparing the results in Figure 9. This figure depicts the Laplace and ordinary periodograms

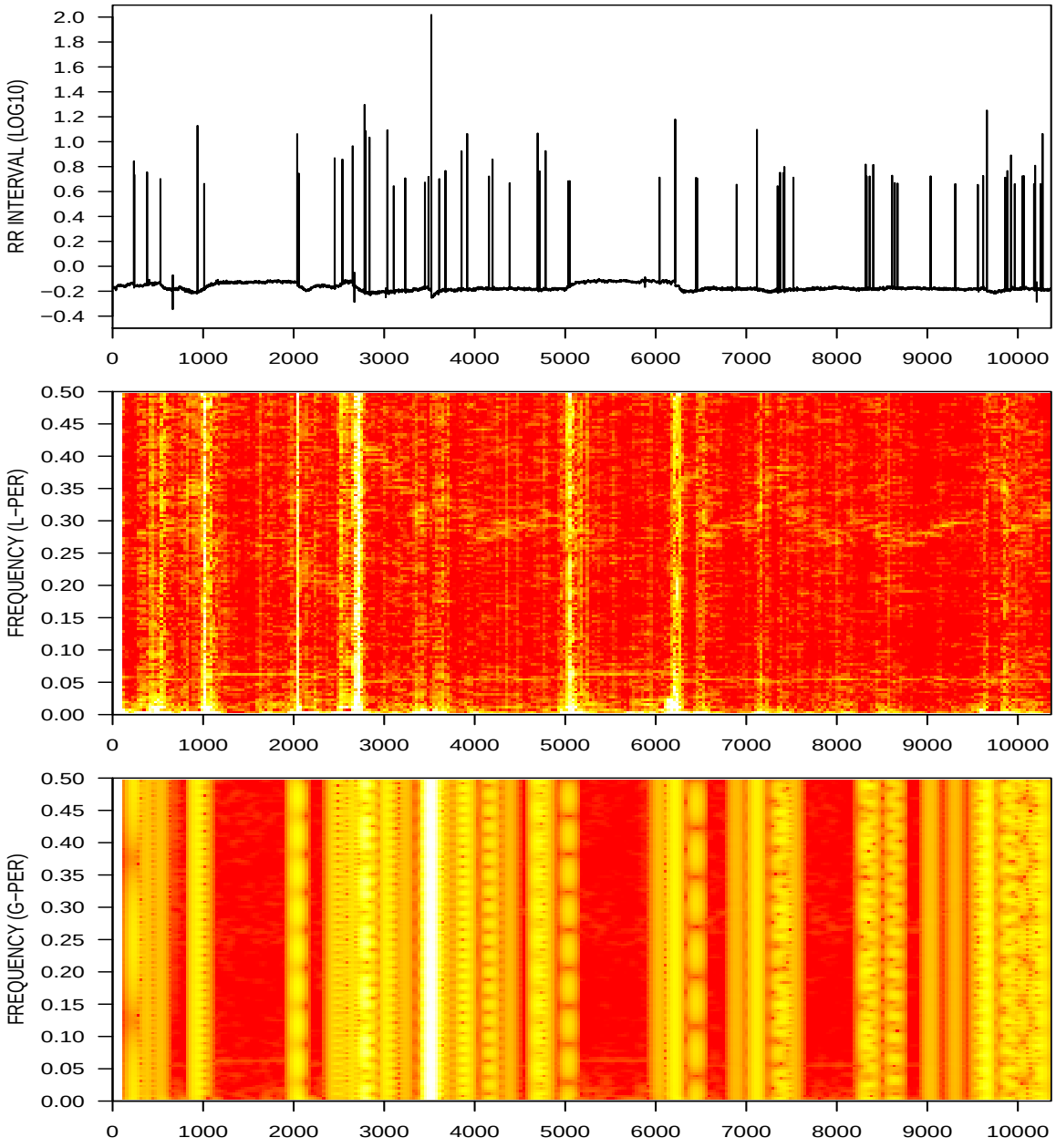


Figure 8: Top, RR tachogram (in log10 of a second); middle, Laplace spectrogram; bottom, ordinary spectrogram. Frequency is measured in cycles per beat.

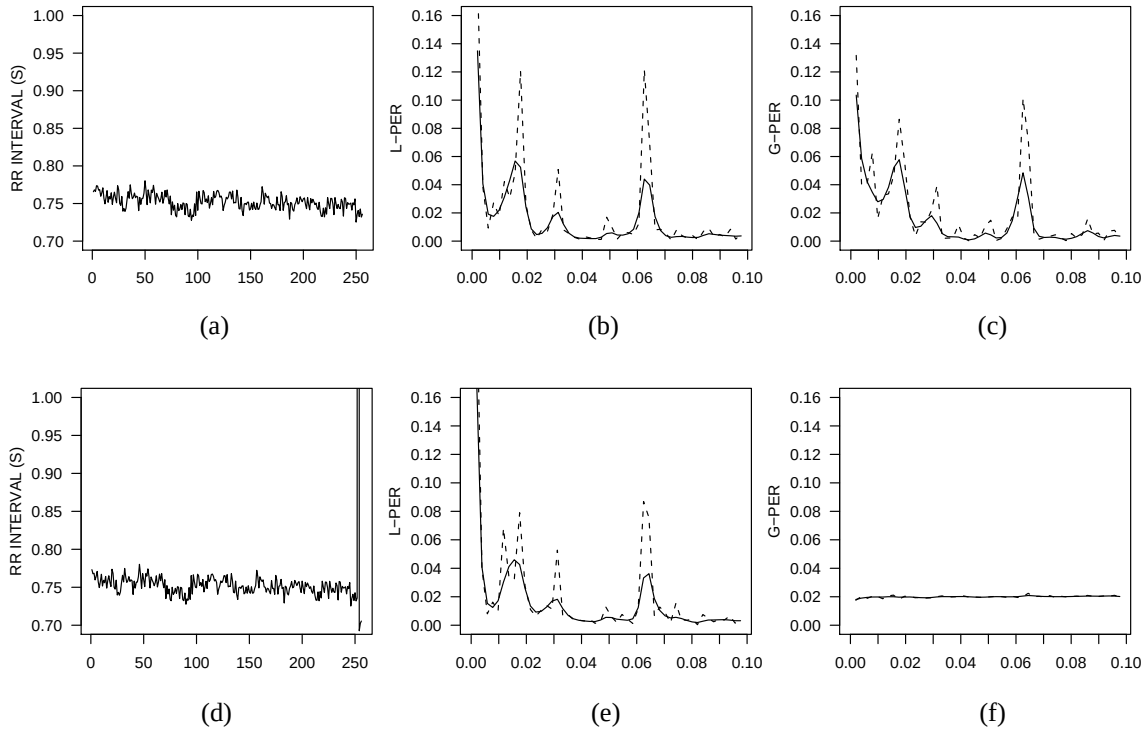


Figure 9: (a)(d) RR tachogram (in seconds). (b)(e) Laplace periodogram (solid line, smoothed; dashed line, raw). (c)(f) Ordinary periodogram.

just before a spike (a–c) and at the onset of the spike (d–f). In the first case, both periodograms exhibit similar spectral patterns (notice the large spectral peak near frequency 0.065). But once a spike occurs, the ordinary periodogram collapses completely, while the Laplace periodogram mostly retains its normal characteristics.

6. CONCLUDING REMARKS

Based on LAD harmonic regression, we have proposed the Laplace periodogram as a counterpart to the ordinary periodogram and investigated its potential as a nonparametric tool for time series analysis. In particular, we have established the fact that the Laplace

periodogram is related to the zero-crossing spectrum in the same way as the ordinary periodogram is related to the autocorrelation spectrum. This relationship justifies the Laplace periodogram for analyzing the serial dependence of time-series data. The theoretical analysis, coupled with the simulation studies, has uncovered the advantage of the Laplace periodogram over the ordinary periodogram in its superior handling of heavy-tailed noise and nonlinear distortion. The real-data example of heart rate variability further confirms the robustness superiority of the Laplace periodogram.

In our experiments, the Laplace periodogram was calculated in R using the `rq` function in the `quantreg` package (Koenker, 2005). It is a general-purpose program designed for linear quantile regression. Although the computational complexity is higher than the Fast Fourier transform, the benefit of improved robustness can still justify the use of the Laplace periodogram in applications where the slower speed of computation is not a primary constraint.

APPENDIX

Proof of Theorem 1

We begin with the single regressor case. We drop the subscripts j and k for simplicity and prove the assertion along the lines of Koenker (2005, pp. 102–122).

First, consider the case where $\{\varepsilon_t\}$ is a sequence of independent random variables, for which $r_{ts}(w_t, w_s) = F_t(w_t)\{1 - F_t(w_t)\}\delta_{t-s}$. For any constant vector $\boldsymbol{\delta} \in \mathbb{R}^p$, let

$$Z_n(\boldsymbol{\delta}) := \tfrac{1}{2} \sum_{t=1}^{n} \{|u_t - v_t(\boldsymbol{\delta})| - |u_t|\}, \tag{12}$$

where $u_t := y_t - \boldsymbol{x}_t^T\boldsymbol{\beta}_0 = \varepsilon_t - w_t$ and $v_t(\boldsymbol{\delta}) := n^{-1/2}\boldsymbol{x}_t^T\boldsymbol{\delta}$. The first step is to establish

$$Z_n(\boldsymbol{\delta}) = \tilde{Z}_n(\boldsymbol{\delta}) + \mathcal{O}_P(1) \tag{13}$$

for each fixed $\boldsymbol{\delta} \in \mathbb{R}^p$, where $\tilde{Z}_n(\boldsymbol{\delta}) := -\boldsymbol{\delta}^T\boldsymbol{\zeta}_n + \frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{Q}_n\boldsymbol{\delta}$ and $\boldsymbol{\zeta}_n \overset{A}{\sim} N(\boldsymbol{h}_n, \boldsymbol{W}_n)$. Toward that end, we use Knight's identity (Knight 1998),

$$\tfrac{1}{2}(|u-v| - |u|) = -v\,\psi(u) + \int_0^v \phi(u,s)\,ds,$$

where $\psi(u) := \frac{1}{2} - I(u < 0)$ and $\phi(u,s) := I(u \le s) - I(u \le 0)$. With u and v substituted by u_t and $v_t := v_t(\boldsymbol{\delta})$, we can write $Z_n(\boldsymbol{\delta}) = Z_{1n} + Z_{2n}$, where $Z_{1n} := -\boldsymbol{\delta}^T\boldsymbol{\zeta}_n$ and $Z_{2n} := \sum \xi_t$, with $\boldsymbol{\zeta}_n := n^{-1/2}\sum \psi(u_t)\boldsymbol{x}_t$ and $\xi_t := \int_0^{v_t} \phi(u_t,s)\,ds$.

Consider $S_n := \sum z_t$, where $z_t := \psi(u_t)\boldsymbol{x}_t^T\boldsymbol{\delta}$ $(t = 1,\ldots,n)$. Owing to the boundedness of $\{\boldsymbol{x}_t\}$ and the independence of $\{\varepsilon_t\}$, the z_t's are independent and uniformly bounded random variables. It is easy to see that $m_t := E\{\psi(u_t)\} = \frac{1}{2} - F_t(w_t)$ and $\sigma_t^2 := V\{\psi(u_t)\} = F_t(w_t)\{1 - F_t(w_t)\}$. Therefore, $E(S_n) = \sum m_t\boldsymbol{x}_t^T\boldsymbol{\delta} = n^{1/2}\boldsymbol{h}_n^T\boldsymbol{\delta}$ and $V(S_n) = \sum \sigma_t^2(\boldsymbol{x}_t^T\boldsymbol{\delta})^2 = n\boldsymbol{\delta}^T\boldsymbol{W}_n\boldsymbol{\delta}$. Under Assumption (iii), we have $V(S_n) \to \infty$. This result, together with the boundedness of $\{z_t\}$, implies that the Lindeberg condition is satisfied. By the central limit theorem (Chung 2001, Theorem 7.2.1), we obtain $S_n \overset{A}{\sim} N(n^{1/2}\boldsymbol{h}_n^T\boldsymbol{\delta}, n\boldsymbol{\delta}^T\boldsymbol{W}_n\boldsymbol{\delta})$, which, in turn, gives $\boldsymbol{\delta}^T\boldsymbol{\zeta}_n = n^{-1/2}S_n \overset{A}{\sim} N(\boldsymbol{h}_n^T\boldsymbol{\delta}, \boldsymbol{\delta}^T\boldsymbol{W}_n\boldsymbol{\delta})$. Because this expression is true for any $\boldsymbol{\delta} \in \mathbb{R}^p$, the Cramér-Wold device (Brockwell and Davis 1992) ensures $\boldsymbol{\zeta}_n \overset{A}{\sim} N(\boldsymbol{h}_n, \boldsymbol{W}_n)$.

Now consider Z_{2n}. Owing to the boundedness of $\{\boldsymbol{x}_t\}$, we have $\max|v_t| = \mathcal{O}(n^{-1/2})$ and

$\sum \mathcal{O}(v_t^{d+2}) = \mathcal{O}(n^{-d/2})$. Therefore, under Assumption (ii),

$$\begin{aligned} E(Z_{2n}) &= \sum_{t=1}^{n} E(\xi_t) = \sum_{t=1}^{n} \int_0^{v_t} \{F_t(u+w_t) - F_t(w_t)\}\, du \\ &= \sum_{t=1}^{n} \int_0^{v_t} \{f_t(w_t)u + \mathcal{O}(u^{d+1})\}\, du = \sum_{t=1}^{n} \{\tfrac{1}{2} f_t(w_t) v_t^2 + \mathcal{O}(v_t^{d+2})\} \\ &= \tfrac{1}{2} \boldsymbol{\delta}^T \boldsymbol{Q}_n \boldsymbol{\delta} + \mathcal{O}(n^{-d/2}). \end{aligned} \tag{14}$$

Furthermore, for $v_t < 0$, $\xi_t = \int_0^{|v_t|} -\phi(u_t, -s)\, ds$, and for $v_t \geq 0$, $\xi_t = \int_0^{v_t} \phi(u_t, s)\, ds$. In both cases, the integrand is bounded by 0 and 1. This implies $0 \leq \xi_t \leq |v_t|$, which, in turn, leads to $V(\xi_t) \leq E(\xi_t^2) \leq |v_t| E(\xi_t)$. Therefore,

$$V(Z_{2n}) = \sum_{t=1}^{n} V(\xi_t) \leq \max_{1 \leq t \leq n} |v_t| E(Z_{2n}) = \mathcal{O}(n^{-1/2}), \tag{15}$$

where the last equality is due to $E(Z_{2n}) = \mathcal{O}(1)$, which is implied by (14) together with Assumption (i) and the boundedness of $\{\boldsymbol{x}_t\}$. From (14) and (15), and by Chebyshev's inequality, we obtain $Z_{2n} = \frac{1}{2} \boldsymbol{\delta}^T \boldsymbol{Q}_n \boldsymbol{\delta} + \mathcal{o}_P(1)$.

Combining the above results proves (13) for each fixed $\boldsymbol{\delta} \in \mathbb{R}^p$. Since $Z_n(\boldsymbol{\delta})$ and $\tilde{Z}_n(\boldsymbol{\delta})$ are convex functions of $\boldsymbol{\delta}$, the Convexity Lemma (Pollard 1991) ensures that (13) also holds uniformly in $\boldsymbol{\delta} \in \Delta$ for any given compact set $\Delta \in \mathbb{R}^p$. Moreover, under Assumption (iv), $\tilde{Z}_n(\boldsymbol{\delta})$ has a unique minimizer $\tilde{\boldsymbol{\delta}}_n := \boldsymbol{Q}_n^{-1} \boldsymbol{\zeta}_n$ with $\boldsymbol{Q}_n \geq \boldsymbol{Q}_0$ for $n \geq N_0$. By using the convexity of $Z_n(\boldsymbol{\delta})$ and following the last part of the proof of Theorem 1 in Pollard (1991), we can show that the minimizer of $Z_n(\boldsymbol{\delta})$, denoted by $\hat{\boldsymbol{\delta}}_n$, is $\mathcal{o}_p(1)$ away from $\tilde{\boldsymbol{\delta}}_n$, i.e., $\hat{\boldsymbol{\delta}}_n - \tilde{\boldsymbol{\delta}}_n \xrightarrow{P} 0$, and therefore has the same asymptotic distribution as $\tilde{\boldsymbol{\delta}}_n$. This result can also be justified by citing the Basic Corollary of Hjort and Pollard (1993). Note that $Z_n(\boldsymbol{\delta})$ can be reparameterized as a function of $\boldsymbol{\beta} := \boldsymbol{\beta}_0 + \boldsymbol{\delta} n^{-1/2}$ such that $Z_n(\boldsymbol{\delta}) = \frac{1}{2} \sum \{|y_t - \boldsymbol{x}_t^T \boldsymbol{\beta}| - |u_t|\}$. Since this function of $\boldsymbol{\beta}$ is minimized at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_n$, it follows that $\hat{\boldsymbol{\delta}}_n = n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$. The proof for the independent

case is complete upon noting that $\tilde{\boldsymbol{\delta}}_n \overset{A}{\sim} N(\boldsymbol{Q}_n^{-1}\boldsymbol{h}_n, \boldsymbol{Q}_n^{-1}\boldsymbol{W}_n\boldsymbol{Q}_n^{-1})$.

Next, consider the case where $\{\varepsilon_t\}$ is an m-dependent process. The proof is similar to that of the independent case, but with several important exceptions. First, $\{z_t\}$ is now an m-dependent and uniformly bounded random process and

$$\sigma_{ts} := \mathrm{Cov}\{\psi(u_t), \psi(u_s)\} = F_{ts}(w_t, w_s) - F_t(w_t)F_s(w_s) = r_{ts}(w_t, w_s).$$

Under Assumption (iii), $V(S_n) = n\boldsymbol{\delta}^T\boldsymbol{W}_n\boldsymbol{\delta}$, so $n^{-2/3}V(S_n) = n^{1/3}\boldsymbol{\delta}^T\boldsymbol{W}_n\boldsymbol{\delta} \to \infty$. By the central limit theorem for uniformly bounded m-dependent processes (Chung 2001, Theorem 7.3.1), we obtain $\boldsymbol{\zeta}_n \overset{A}{\sim} N(\boldsymbol{h}_n, \boldsymbol{W}_n)$. Second, since $V(\xi_t) \le |v_t|E(\xi_t)$, we have

$$|\mathrm{Cov}(\xi_t, \xi_s)| \le \sqrt{V(\xi_t)V(\xi_s)} \le \sqrt{|v_t v_s|E(\xi_t)E(\xi_s)}.$$

In addition, the m-dependence of $\{u_t\}$ gives $\mathrm{Cov}(\xi_t, \xi_s) = 0$ for $|t-s| > m$. Combining these results with the Cauchy-Schwartz inequality leads to

$$\begin{aligned} V(Z_{2n}) &= \sum_{|t-s|\le m} \mathrm{Cov}(\xi_t, \xi_s) \le \sum_{|t-s|\le m} \sqrt{|v_t v_s|E(\xi_t)E(\xi_s)} \\ &\le \sum_{|t-s|\le m} |v_t|E(\xi_t) \le \sum_{t=1}^{n}(2m+1)|v_t|E(\xi_t) \\ &\le (2m+1)\max_{1\le t\le n}|v_t|E(Z_{2n}) = \mathcal{O}(mn^{-1/2}) = \mathcal{o}(1). \end{aligned}$$

The remaining argument is the same as in the proof of the independent case.

Finally, consider the case of multiple regressors. With loss of generality, let $q = 2$. To prove the joint asymptotic normality of $\hat{\boldsymbol{\beta}}_{1n}$ and $\hat{\boldsymbol{\beta}}_{2n}$, it suffices to consider $Z_n(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) := Z_n^{(1)}(\boldsymbol{\delta}_1) + Z_n^{(2)}(\boldsymbol{\delta}_2)$, where $Z_n^{(j)}(\boldsymbol{\delta}_j)$ is defined in the same as $Z_n(\boldsymbol{\delta})$ in (12) except that $\boldsymbol{x}_{jt}$ is in place of $\boldsymbol{x}_t$. It has been shown that $Z_n^{(j)}(\boldsymbol{\delta}_j) = -\boldsymbol{\delta}_j^T\boldsymbol{\zeta}_{jn} + \boldsymbol{\delta}_j^T\boldsymbol{Q}_{jn}\boldsymbol{\delta}_j + \mathcal{o}_P(1)$ uniformly in $\boldsymbol{\delta}_j \in \Delta$, where $\boldsymbol{\zeta}_{jn} \overset{A}{\sim} N(\boldsymbol{h}_{jn}, \boldsymbol{W}_{jjn})$. Therefore, we only need to compute $\mathrm{Cov}(\boldsymbol{\zeta}_{1n}, \boldsymbol{\zeta}_{2n})$.

By definition, $\boldsymbol{\zeta}_{jn} = n^{-1/2}\sum \psi(u_{jt})\boldsymbol{x}_{jt}$, where $u_{jt} := \varepsilon_t - w_{jt}$. Since $\mathrm{Cov}\{\psi(u_{jt}), \psi(u_{ks})\} = r_{ts}(w_{jt}, w_{ks})$, we obtain $\mathrm{Cov}(\boldsymbol{\zeta}_{1n}, \boldsymbol{\zeta}_{2n}) = n^{-1}\sum\sum r_{ts}(w_{1t}, w_{2s})\boldsymbol{x}_{1t}\boldsymbol{x}_{2s}^T = \boldsymbol{W}_{12n}$. □

Remark 4. A strengthened version of Theorem 7.3.1 (Chung 2001) states that if $\{X_t\}$ $(t = 1, \ldots, n)$ is a sequence of m-dependent and uniformly bounded random variables with $m = \mathcal{O}(n^{\delta})$ and $n^{-\varepsilon}V(S_n) \to \infty$ for some $\delta \in [0, 1)$ and $\varepsilon := \frac{2}{3}(1 + 2\delta)$, where $S_n := \sum X_t$, then $V(S_n)^{-1/2}(S_n - E(S_n)) \xrightarrow{D} N(0, 1)$. This assertion can be proved in the same way as in Chung (2001, pp. 224–226) by taking the blocking factor $k := [n^{\varepsilon - 2\delta}]$ so that $km^2 = \mathcal{O}(n^{\varepsilon})$, $k = \mathcal{O}(n^{\varepsilon - 2\delta}) \to \infty$, $n/(km) = \mathcal{O}(n^{1-\varepsilon+\delta}) \to \infty$, and $n/k = \mathcal{O}(n^{1-\varepsilon+2\delta}) = \mathcal{O}(n^{\varepsilon/2})$. Citing this result, we can show that the assertion in Theorem 1 remains valid for m-dependent processes with $m = \mathcal{O}(n^{\delta})$ for some $\delta \in [0, \frac{1}{4})$. Indeed, since $\varepsilon := \frac{2}{3}(1 + 2\delta) < 1$, by following the proof of Theorem 1, we obtain $n^{-\varepsilon}V(S_n) = n^{1-\varepsilon}\boldsymbol{\delta}^T\boldsymbol{W}_n\boldsymbol{\delta} \to \infty$. We also have $V(Z_{2n}) = \mathcal{o}(1)$ because $mn^{-1/2} = \mathcal{O}(n^{\delta - 1/2}) \to 0$. Combining these modifications with the rest of the argument in the proof of Theorem 1 proves the assertion.

Proof of Corollary 1

According to the proof of Theorem 1, it suffices to establish the quadratic approximation (13). Towards that end, we note that ε_t can be decomposed as $\varepsilon_t = \varepsilon_t' + \varepsilon_t''$, where

$$\varepsilon_t' := \sum_{|j| \le m} \phi_j e_{t-j}, \quad \varepsilon_t'' := \sum_{|j| > m} \phi_j e_{t-j}.$$

With this notation, we can write $u_t := \varepsilon_t - w_t = u_t' + \varepsilon_t''$, where $u_t' := \varepsilon_t' - w_t$. We can also write $Z_n(\boldsymbol{\delta}) = Z_n'(\boldsymbol{\delta}) + Z_n''(\boldsymbol{\delta})$, where $Z_n(\boldsymbol{\delta})$ is defined by (12) and $Z_n'(\boldsymbol{\delta}) := \frac{1}{2}\sum(|u_t' - v_t| - |u_t'|)$. Note that $\{\varepsilon_t'\}$ is a $2m$-dependent process. Therefore, according to Remark 4, a quadratic approximation of the form (13) can be obtained for $Z_n'(\boldsymbol{\delta})$ with $m = \mathcal{O}(n^{\delta})$ for any $\delta \in [0, \frac{1}{4})$,

where the parameters of the quadratic function are defined by the marginal and bivariate distribution functions of $\{\varepsilon_t\}$ as the limiting values of the corresponding functions of $\{\varepsilon'_t\}$. Given this result, it suffices to show that $Z''_n(\boldsymbol{\delta}) = \mathcal{O}_P(1)$.

Towards that end, we note that $Z''_n(\boldsymbol{\delta}) = Z''_{1n} + Z''_{2n}$, where

$$Z''_{1n} := \tfrac{1}{2}\sum_{t=1}^{n}(|u''_t + \varepsilon''_t| - |u''_t|), \quad Z''_{2n} := -\tfrac{1}{2}\sum_{t=1}^{n}(|u'_t + \varepsilon''_t| - |u'_t|),$$

and $u''_t := u'_t - v_t$. Using the Knight's identity, we can show that $|Z''_{1n}| \leq 2\sum|\varepsilon''_t|$. This, combined with $|\varepsilon''_t| \leq \sum_{|j|>m}|\phi_j|\,|e_{t-j}|$, leads to

$$E(|Z''_{1n}|) \leq 2\sum_{t=1}^{n}\sum_{|j|>m}|\phi_j|E(|e_{t-j}|) = 2nE(|e_0|)\sum_{|j|>m}|\phi_j| = \mathcal{O}(1).$$

Similarly, $E(|Z''_{2n}|) = \mathcal{O}(1)$. Citing Markov's inequality completes the proof. □

Proof of Theorem 2

In Theorem 1, let $\boldsymbol{x}_{jt} := \boldsymbol{x}_t(\omega_j)$, $\mu_t = 0$, and $\boldsymbol{\beta}_0 = \mathbf{0}$. Then, the white noise assumption implies that $\boldsymbol{W}_{jkn} = \frac{1}{4}D_{jkn}$ and $\boldsymbol{Q}_{jn} = f(0)\boldsymbol{D}_{jjn}$. This, combined with Assumption (vii), proves the assertion. □

Proof of Theorem 3

First, consider the single frequency case. For simplicity, we drop the subscripts j and k in the notation. In Theorem 1, let $\boldsymbol{x}_t := \boldsymbol{x}_t(\omega)$ and $\boldsymbol{\beta}_0 = \mathbf{0}$ so that $w_t = -\mu_t = 0$. Since $F(0) = \frac{1}{2}$, we have $r_{ts} = \frac{1}{4} - \frac{1}{2}\gamma_{ts} = \frac{1}{4} - \frac{1}{2}\gamma_{t-s} := r_{t-s}$. This implies that

$$\boldsymbol{W}_n = n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} r_{t-s}\boldsymbol{x}_t\boldsymbol{x}_s^T = n^{-1}\sum_{|\tau|<n} r_\tau \sum_{t\in T_n(\tau)} \boldsymbol{x}_t\boldsymbol{x}_{t-\tau}^T, \tag{16}$$

where $T_n(\tau) := \{t : \max(1, 1+\tau) \le t \le \min(n, n+\tau)\}$. Since $\boldsymbol{x}_t = \boldsymbol{x}_t(\omega)$, we have

$$\boldsymbol{x}_t \boldsymbol{x}_{t-\tau}^T = \begin{bmatrix} c_\tau c_t^2 + s_\tau c_t s_t & c_\tau c_t s_t - s_\tau c_t^2 \\ c_\tau c_t s_t + s_\tau s_t^2 & c_\tau s_t^2 - s_\tau c_t s_t \end{bmatrix},$$

where $c_t := \cos(\omega t)$ and $s_t := \sin(\omega t)$. It is easy to show that $\sum c_t^2 = \frac{1}{2}n + \mathcal{O}(1)$, $\sum s_t^2 = \frac{1}{2}n + \mathcal{O}(1)$, and $\sum c_t s_t = \mathcal{O}(1)$. Therefore, for any fixed τ,

$$n^{-1} \sum_{t \in T_n(\tau)} \boldsymbol{x}_t \boldsymbol{x}_{t-\tau}^T \to \boldsymbol{M}_\tau := \tfrac{1}{2} \begin{bmatrix} c_\tau & -s_\tau \\ s_\tau & c_\tau \end{bmatrix}.$$

Moreover, since $\{r_\tau\}$ is absolutely summable, we obtain $\boldsymbol{W}_n \to \sum r_\tau \boldsymbol{M}_\tau$. Because r_τ is an even function of τ, we have $\sum r_\tau s_\tau = 0$; because $r_\tau = \frac{1}{4}(1 - 2\gamma_\tau)$, we have $\sum r_\tau c_\tau = \frac{1}{4}S(\omega)$. Combining these results yields $\boldsymbol{W}_n \to \frac{1}{8}S(\omega)\mathbf{I}$. Moreover, we also have $\boldsymbol{h}_n = \mathbf{0}$ and $\boldsymbol{Q}_n \to \frac{1}{2}f(0)\mathbf{I}$. Therefore, by Theorem 1 and Corollary 1, $n^{1/2}\hat{\boldsymbol{\beta}}_n(\omega) \overset{A}{\sim} N(\mathbf{0}, \boldsymbol{Q}_n^{-1}\boldsymbol{W}_n\boldsymbol{Q}_n^{-1}) \to N(\mathbf{0}, (1/2f^2(0))S(\omega)\mathbf{I})$.

Now, consider the case of multiple frequencies. It suffices to show $\boldsymbol{W}_{jkn} = \mathcal{o}(1)$ for $j \neq k$. Toward that end, we note that $\boldsymbol{W}_{jkn}$ has the same expression as $\boldsymbol{W}_n$ in (16) except that $\boldsymbol{x}_t(\omega_j)\boldsymbol{x}_{t-\tau}^T(\omega_k)$ is in place of $\boldsymbol{x}_t\boldsymbol{x}_{t-\tau}^T$. For $j \neq k$, Assumption (vii) implies that $n^{-1}\sum_{t \in T_n(\tau)} \boldsymbol{x}_t(\omega_j)\boldsymbol{x}_{t-\tau}^T(\omega_k) = \mathcal{o}(1)$ for each fixed τ, which, combined with the absolute summability of $\{r_\tau\}$, leads to $\boldsymbol{W}_{jkn} = \mathcal{o}(1)$. □

Proof of Theorem 4

In Theorem 1 (with $q = 1$ and with subscripts j and k omitted), let $\boldsymbol{x}_t := \boldsymbol{x}_t(\omega_0)$ and $w_t = 0$. Then, we have $\hat{\boldsymbol{\beta}}_n(\omega_0) \overset{A}{\sim} N(\boldsymbol{\beta}_0, n^{-1}\boldsymbol{Q}_n^{-1}\boldsymbol{W}_n\boldsymbol{Q}_n^{-1})$. The proof of Theorem 3 establishes the fact that $\boldsymbol{Q}_n \to \frac{1}{2}f(0)\mathbf{I}$ and $\boldsymbol{W}_n \to \frac{1}{8}S(\omega_0)\mathbf{I}$. Combining these results proves the assertion. □

Proof of Theorem 5

The asymptotic normality of $\hat{\boldsymbol{\beta}}_n(\omega)$ is a direct result of Theorem 1 and the fact that $b_0 > 0$. The asymptotic distribution of $L_n(\omega)$ is derived from the theory of quadratic forms of multivariate normal random variables. □

Proof of Equation (7)

First, since $\frac{1}{2} - F(w_t)$ is a periodic sequence of period $p := 2\pi/\omega_0 = n/k_0 = 2m+1 > 2$, we have the discrete Fourier transform (DFT) representation $\frac{1}{2} - F(w_t) = a_0 + \sum_{\ell=1}^{m} \boldsymbol{a}_\ell^T \boldsymbol{x}_t(\omega'_\ell)$, where $\omega'_\ell := 2\pi\ell/p = 2\pi\ell k_0/n$. Let $c_t := \cos(\omega t)$, $s_t := \sin(\omega t)$, $c'_t := \cos(\omega'_\ell t)$, and $s'_t := \sin(\omega'_\ell t)$. Since $\sum c_t c'_t = \sum s_t s'_t = \frac{1}{2} n \delta_{\ell k_0 - k}$ and $\sum c_t s'_t = \sum s_t c'_t = 0$, we have

$$n^{-1} \sum_{t=1}^{n} \{\tfrac{1}{2} - F(w_t)\} \boldsymbol{x}_t = \tfrac{1}{2} \boldsymbol{a}. \tag{17}$$

Since $f(w_t)$ is also periodic with period p, we have the DFT representation $f(w_t) = b_0 + \sum_{\ell=1}^{m} \boldsymbol{b}_\ell^T \boldsymbol{x}_t(\omega'_\ell)$. Using the fact that $\sum c'_t c_t^2 = -\sum c'_t s_t^2 = \sum s'_t c_t s_t = \frac{1}{4} \left(\delta_{\ell k_0 - 2k} + \delta_{n - \ell k_0 - 2k}\right)$ and $\sum c'_t c_t s_t = \sum s'_t c_t^2 = \sum s'_t s_t^2 = 0$ and with the notation

$$\boldsymbol{B}_\ell := \frac{1}{2b_0} \begin{bmatrix} b_{1\ell} & b_{2\ell} \\ b_{2\ell} & -b_{1\ell} \end{bmatrix}, \quad \boldsymbol{b}_\ell := [b_{1\ell}, b_{2\ell}]^T,$$

we obtain

$$\boldsymbol{Q}_n = \tfrac{1}{2} b_0 \boldsymbol{B}. \tag{18}$$

Similarly, using the DFT representation $F(w_t)\{1 - F(w_t)\} = c_0 + \sum_{\ell=1}^{m} \boldsymbol{c}_\ell^T \boldsymbol{x}_t(\omega'_\ell)$ with

$$\boldsymbol{C}_\ell := \frac{1}{2c_0} \begin{bmatrix} c_{1\ell} & c_{2\ell} \\ c_{2\ell} & -c_{1\ell} \end{bmatrix}, \quad \boldsymbol{c}_\ell := [c_{1\ell}, c_{2\ell}]^T,$$

we obtain $\boldsymbol{W}_n = \frac{1}{2} c_0 \boldsymbol{C}$. Combining this result with (17) and (18) proves (7). □

REFERENCES

Acar, B., Savelieva, I., Hemingway, H., and Malik, M. (2000), "Automatic ectopic beat elimination in short-term heart rate variability measurement," *Computer Methods and Programs in Biomedicine*, 63, 123–131.

Albrecht, P. and Cohen, R. J., (1988), "Estimation of heart rate power spectrum bands from real-world data: dealing with ectopic beats and noisy data," *Proceedings of Computers in Cardiology*, (Washington, DC, September 25–28, 1988), 311–314.

Arcones, M. A. (2001), "Asymptotic distribution of regression M-estimators," *Journal of Statistical Planning and Inference*, 97, 235–261.

Bahai, A. R. S., Singh, M., Goldsmith, A. J., and Saltzberg, B. R. (2002), "A new approach for evaluating clipping distortion in multicarrier systems," *IEEE Journal on Selected Areas in Communications*, 20, 1037–1046.

Bantli, F., and Hallin, M. (1999), "L_1-estimation in linear models with heterogeneous white noise," *Statistics & Probability Letters*, 45, 305–315.

Bloomfield, P., and Steiger, W. L. (1983), *Least Absolute Deviations*, Boston, MA: Birkhäuser.

Breidt, F. J., Davis, R. A., and Trindate, A. A. (2001), "Least absolute deviation estimation for all-pass time series models," *Annals of Statistics*, 29, 919–946.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd Edition, New York: Springer.

Chorti, A., and Brookes, M. (2006), "On the effects of memoryless nonlinearities on M-QAM and DQPSK OFDM signals," *IEEE Transactions on Microwave Theory and Techniques*, 54, 3301–3315.

Chung, K. L. (2001), *A Course in Probability Theory*, 3rd Edition, New York: Academic Press.

Clifford, G. D., and Tarassenko, L. (2005), "Quantifying errors in spectral estimates of HRV due to beat replacement and resampling," *IEEE Transactions on Biomedical Engineering*, 52, 630–638.

Davis, R. A., and Dunsmuir, W. T. M. (1997), "Least absolute deviation estimation for regression with ARMA errors," *Journal of Theoretical Probability*, 10, 481–497.

Dielman, T. E. (2005), "Least absolute value regression: recent contributions," *Journal of Statistical Computation and Simulation*, 75, 263–286.

Dodge, Y., Ed., (1997), *L_1-Statistical Procedures and Related Topics*, Institute of Mathematical Statistics Lecture Notes-Monograph Series, vol. 31.

Dodge, Y., Ed., (2002), *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, Boston, MA: Birkhäuser.

Hjort, N. L., and Pollard, D., (1993), "Asymptotics for minimisers of convex processes," Tech. Report, Dept. of Statistics, Yale Univ., New Heaven, CT. Available at www.stat.yale.edu/preprints/1993/93may-1.pdf.

Kedem, B. (1994), *Time Series Analysis by Higher Order Crossings*, Piscataway, NJ: IEEE Press.

Knight, K. (1998), "Limiting distributions for L_1 regression estimators under general conditions," *Annals of Statistics*, 26, 755–770.

Koenker, R. (2005), *Quantile Regression*, Cambridge, UK: Cambridge University Press.

Lippman, N., Stein, K. M., and Lerman, B. B. (1994), "Comparison of methods for removal of ectopy in measurement of heart rate variability," *American Journal of Physiology – Heart and Circulatory Physiology*, 267, 411–418.

Lai, P. Y., and Lee, S. M. S. (2005), "An overview of asymptotic properties of L_p regression under general classes of error distributions," *Journal of the American Statistical Association*, 100, 446–458.

Malik, M., and Camm, A. J., (eds), (1995), *Heart Rate Variability*, Armonk, NY: Futura Publishing Company.

Malik, M., et al. (1996), "Heart rate variability: standards of measurement, physiological interpretation and clinical use," Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, *Circulation*, 93, 1043–1065.

Mateo, J., and Laguna, P. (2003), “Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal,” *IEEE Transactions on Biomedical Engineering*, 50, 334–343.

Pollard, D. (1991), “Asymptotics for least absolute deviation regression estimators,” *Econometric Theory*, 7, 186–199.

Portnoy, S. (1991), “Asymptotic behavior of regression quantiles in non-stationary, dependent cases,” *Journal of Multivariate Analysis*, 38, 100–113.

Quinn, B. G., and Hannan, E. J. (2001), *The Estimation and Tracking of Frequency*, Cambridge, UK: Cambridge University Press.

Storck, N., Ericson, M., Lindblad, L. E., and Jensen-Urstad, M. (2001), “Automatic computerized analysis of heart rate variability with digital filtering of ectopic beats,” *Clinical Physiology*, 21, 15–24.

Weiss, A. A. (1990), “Least absolute error estimation in the presence of serial correlation,” *Journal of Econometrics*, 44, 127–158.

Wise, G., Traganitis, A., and Thomas, J. (1977), “The effect of a memoryless nonlinearity on the spectrum of a random process,” *IEEE Transactions on Information Theory*, 23, 84–89.

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