Expectile Periodogram

Tianbo Chen [*] Ta-hsin Li [†]

Abstract

This paper introduces a novel periodogram-like function, called the expectile periodogram, for modeling spectral features of time series and detecting hidden periodicities. The expectile periodogram is constructed from trigonometric expectile regression, in which a specially designed check function is used to substitute the squared l_2 norm that leads to the ordinary periodogram. The expectile periodogram retains the key properties of the ordinary periodogram as a frequency-domain representation of serial dependence in time series, while offering a more comprehensive understanding by examining the data across the entire range of expectile levels. We establish the asymptotic theory and investigate the relationship between the expectile periodogram and the so-called expectile spectrum. Simulations demonstrate the efficiency of the expectile periodogram in the presence of hidden periodicities. Finally, by leveraging the inherent two-dimensional nature of the expectile periodogram, we train a deep learning (DL) model to classify earthquake waveform data. Remarkably, our approach outperforms alternative periodogram-based methods in terms of classification accuracy.

Key Words: Periodogram; Expectile regression; Expectile spectrum; Time series analysis; Earthquake data classification

1 Introduction

Spectral analysis play a crucial role in time series analysis, where data are analyzed in the frequency domain. The ordinary periodogram (PG), a raw non-parametric estima-

[*]Anhui University, Hefei, China. E-mail: chentianbo@ahu.edu.cn
[†]IBM Watson Research Center, NY, US. E-mail: thl@us.ibm.com.

tor of the power spectrum, is widely applied across various fields (Caiado et al., 2006; Polat and Güneş, 2007; Baud et al., 2018; Euán et al., 2018; Maadooliat et al., 2018; Martínez-Murcia et al., 2019; Chen et al., 2021a). The PG is constructed by the ordinary least squares (OLS) regression on the trigonometric regressor. However, OLS exhibits limitations in terms of robustness and effectiveness, particularly in handling data with asymmetric or heavy-tailed distributions. Moreover, OLS focuses primarily on the conditional mean, lacking the ability to provide a comprehensive view of the data, which subsequently affects the performance of the PG (Bloomfield, 2004).

An alternative regression approach is quantile regression, which evaluates the variation of conditional quantiles with respect to the response variable. The pioneering work of Koenker and Bassett Jr (1978) introduced the concept of quantile regression, and it has been comprehensively extended by Koenker (2005). For a detailed and systematic introduction to quantile regression and its extensions, we refer to Kouretas et al. (2005); Cai and Xu (2008); Cai and Xiao (2012); Koenker (2017). Equipped with a specially designed check function, quantile regression provides a more complete picture of the relationship between the response variable and the covariates, while also demonstrating strong robustness against outliers. As a result, quantile regression has been applied in various fields (Garcia et al., 2001; Machado and Mata, 2005; Alvarado et al., 2021; Sharif et al., 2021). One innovative development of quantile regression is the quantile periodogram (QP) (Li, 2012b) constructed by trigonometric quantile regression, which demonstrates its ability to detect hidden periodicities in time series. Similarly to the behavior of the power spectrum and the PG, the QP is an unbiased estimator of the so-called quantile spectrum, which is a scaled version of the ordinary power spectrum of the level-crossing process. Notably, the Laplace periodogram (Li, 2008) represents a specialized case of the QP when the quantile is set to 0.5. Related works on the QP include Li (2012a); Hagemann (2013); Li (2014); Dette et al. (2015); Kley (2016); Birr et al. (2017); Meziani et al. (2020); Chen et al. (2021b); Li (2023).

However, quantile regression is beset by the computational burden due to the non-differentiability of its check function. To address the issue and strike a balance between robustness and effectiveness, expectile regression has been proposed. Asymmetric least square (ALS) regression, also known as expectile regression, was proposed by Newey and Powell (1987). Expectile regression can capture the entire distribution of the data and is much easier to compute using weighted least squares estimators. Furthermore, consistent estimation of the joint asymptotic covariance matrix of several ALS estimators does not require estimation of the density function of the error terms. A comprehensive comparative analysis of quantiles and expectiles is presented in Waltrup et al. (2015), wherein the relationships between these two approaches are thoroughly examined. In addition, Jones (1994) provided a mathematical proof that expectiles indeed correspond to quantiles of a distribution function uniquely associated with the underlying data distribution. Building on this foundation, Yao and Tong (1996) established the existence of a bijective function that directly relates expectiles to quantiles, thereby facilitating the calculation of one from the other. Alternative approaches for estimating quantiles from expectiles are introduced in Efron (1991); Granger and Sin (1997); Schnabel and PHC (2009), offering methodologies for estimating the density (and also, quantiles) from a set of expectiles by using penalized least squares. As a generalization of both quantile and expectile regression, Jiang et al. (2021) introduced the k-th power expectile regression with $1<k \leq 2$.

Expectile regression techniques are applied across diverse domains. For instance, Jiang et al. (2017) introduced an expectile regression neural network (ERNN) model, which incorporates a neural network structure into expectile regression, thereby facilitating the exploration of potential nonlinear relationships between covariates and the expectiles of the response variable. Gu and Zou (2016) systematically studied the sparse asymmetric least squares (SALES) regression under high dimensions where the penalty functions include the Lasso and nonconvex penalties. Xu et al. (2020) developed a novel mixed data sampling expectile regression (MIDAS-ER) model to measure financial risk and demonstrated exceptional performance when applied to two popular financial risk measures: value at risk (VaR) and expected shortfall (ES). Another approach by Xu et al. (2021) involved the elastic-net penalty into expectile regression, and applied the model to two real-world applications: relative location of CT slices on the axial axis and metabolism of tacrolimus drug.

Inspired by the notable success of expectile regression and the foundational work of the QP, we propose a novel spectral estimator termed the expectile periodogram, for spectral analysis of time series data. We demonstrates that the expectile periodogram not only shares similar properties of the PG as a frequency-domain representation of serial dependence in time series, but also provides a more comprehensive analysis by exploring the full range of expectile levels.

The remainder of the paper is organized as follows. In Section 2, we define the expectile periodogram and demonstrate its capability in detecting hidden periodicities through two real-world examples. We establish the asymptotic analysis in Section 3. In Section 4, we present comparative studies evaluating the performance of different periodograms by simulations. In Section 5, we apply our method to a time-series classification task, where we leverage the two-dimensional nature of the expectile periodogram and training a deep learning (DL) model to classify the earthquake data. Conclusions and future works are discussed in Section 6.

2 Expectile Periodogram

2.1 Expectile and Expectile Regression

Let $\{y_t\}$ be a stationary process with cumulative distribution function $F(\cdot)$ and finite second moments. For any $\alpha \in (0,1)$, define $\rho_\alpha(u) := u^2|\alpha - I(u<0)|$, i.e.,

$$\rho_\alpha(u) := \begin{cases} \alpha u^2 & \text{if } u \geq 0, \\ (1-\alpha)u^2 & \text{otherwise.} \end{cases}$$

The α-expectile of y_t is then defined as

$$\mu(\alpha) := \arg\min_{\mu \in \mathbb{R}} \mathrm{E}\left\{\rho_\alpha(y_t - \mu)\right\}, \tag{1}$$

which satisfies the normal equation

$$\mathrm{E}\left\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\right\} = 0. \tag{2}$$

Given a time series $\{y_t : t = 1, \ldots, n\}$ of length n and a deterministic sequence of regressors $\{\mathbf{x}_{nt} : t = 1, \ldots, n\}$ with $\mathbf{x}_{nt} := [x_{nt}(1), \ldots, x_{nt}(p)]^T$ and $x_{nt}(1) := 1$, the expectile regression (ER) at level α solves the optimization problem:

$$\widehat{\boldsymbol{\beta}}_n(\alpha) := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{t=1}^{n} \rho_\alpha(y_t - \mathbf{x}_{nt}^T \boldsymbol{\beta}). \tag{3}$$

In the special case of $\alpha = 0.5$, the ER problem (3) reduces to the ordinary least squares regression because $\rho_{0.5}(u) = 0.5u^2$ for all $u \in \mathbb{R}$. In the special case of $p = 1$, the ER solution $\widehat{\beta}_n(\alpha) := \arg\min_{\beta\in\mathbb{R}} \sum_{t=1}^n \rho_\alpha(y_t - \beta)$ is called the sample α-expectile and is a natural estimator of $\mu(\alpha)$.

2.2 Expectile Discrete Fourier Transform

Let $\omega_\nu := 2\pi\nu/n$ for $\nu = 0, 1, \ldots, n-1$ denote a Fourier frequency for a time series of length n. For $\omega_\nu \notin \{0, \pi\}$, let $\mathbf{x}_{nt}$ in (3) be a trigonometric regressor of the form

$$\mathbf{x}_t(\omega_\nu) := [1, \cos(\omega_\nu t), \sin(\omega_\nu t)]^T, \tag{4}$$

and denote the resulting ER solution by

$$\widehat{\boldsymbol{\beta}}_n(\omega_\nu, \alpha) := [\widehat{\beta}_1(\omega_\nu, \alpha), \widehat{\beta}_2(\omega_\nu, \alpha), \widehat{\beta}_3(\omega_\nu, \alpha)]^T.$$

For $\omega_\nu = \pi$, let $\mathbf{x}_{nt}$ take the form

$$\mathbf{x}_t(\pi) := [1, \cos(\pi t)]^T,$$

and let the resulting ER solution be denoted by

$$\widehat{\boldsymbol{\beta}}_n(\pi, \alpha) := [\widehat{\beta}_1(\pi, \alpha), \widehat{\beta}_2(\pi, \alpha)]^T.$$

Finally, for $\omega_\nu = 0$, let $\widehat{\beta}_1(0, \alpha)$ denote the ER solution with the regressor $\mathbf{x}_{nt} := 1$.

Given these trigonometric ER solutions, the expectile discrete Fourier transform (EDFT) of $\{y_t : t = 1, \ldots, n\}$ at level α is defined as

$$z_n(\omega_\nu, \alpha) := \begin{cases} n\hat{\beta}_1(0, \alpha) & \omega_\nu = 0, \\ n\hat{\beta}_2(\pi, \alpha) & \omega_\nu = \pi \text{ (if } n \text{ is even)}, \\ (n/2)\{\hat{\beta}_2(\omega_\nu, \alpha) - i\hat{\beta}_3(\omega_\nu, \alpha)\} & \text{otherwise,} \end{cases}$$

where $i := \sqrt{-1}$. The EDFT can be viewed as an extension of the ordinary DFT defined by

$$z_n(\omega_\nu) := \sum_{t=1}^n y_t \exp(-i\omega_\nu t),$$

in the sense that the EDFT at $\alpha = 0.5$ coincides with the ordinary DFT, i.e., $z_n(\omega_\nu, 0.5) = z_n(\omega_\nu)$.

2.3 Expectile Periodogram

Given the EDFT $\{z_n(\omega_\nu, \alpha) : \nu = 0, 1, \ldots, n-1\}$, the expectile periodogram (EP) at expectile level α is defined as

$$EP_n(\omega_\nu, \alpha) := n^{-1}|z_n(\omega_\nu, \alpha)|^2 \quad (\nu = 0, 1, \ldots, n-1). \tag{5}$$

In the special case of $\alpha = 0.5$, the EP reduces to the PG:

$$I_n(\omega_\nu) := n^{-1}\left|\sum_{t=1}^{n} y_t \exp(-i\omega_\nu t)\right|^2 = n^{-1}|z_n(\omega_\nu)|^2.$$

Due to symmetry, it suffices to restrict the EP to $\omega_\nu \in [0, \pi]$. Also, it is often convenient to define $EP_n(0, \alpha) = 0$ instead of $n|\hat{\beta}_1(0, \alpha)|^2$. This is analogous to the property of the PG of the demeaned series $\{y_t - \bar{y} : t = 1, \ldots, n\}$, where $\bar{y} := n^{-1}\sum_{t=1}^{n} y_t$.

Remark 1: In Section 2.4, 4, and 5, our primary focus is on analyzing the serial dependence of the time series, with amplitude considerations being secondary. Accordingly, we normalize all the periodograms at each expectile level such that the summation over ω equals 1. Additionally, due to symmetry, we only consider restricted $\omega_\nu \in [0, \pi]$, and use $f = \omega/2\pi \in [0, 1/2]$, which represents the number of cycles per unit time, in all figures of this paper.

2.4 Examples

While the EP can be analyzed at a specific expectile level like the PG, it can also be treated as a bivariate function of parameters α and ω. To illustrate such duality, we consider two examples in different scientific fields, in which the PGs fail to capture full spectral features of the time series.

In the first example, we analyze the daily log returns of the S&P 500 Index data from 1986 to 2015, as shown in Figure 1(a). The EP in Figure 1(b) successfully identifies the approximately 10-year cycle of market volatility (highlighted by the blue lines in Figure 1(a)) at both the lower and upper expectiles. In contrast, the PG yields a relatively featureless flat line. The smoothed periodograms in Figure 1(d) and (e), obtained using the smooth.spline function in R with the tuning parameter cv = TRUE, exhibit a spectral feature similar to a GARCH(1,1) (Engle, 1982; Bollerslev, 1986) model:

$$y_t \sim N(0, \sigma_t^2), \tag{6}$$

where $\sigma_t^2 = 10^{-6} + 0.49y_{t-1}^2 + 0.49\sigma_{t-1}^2$. We generate 5,000 realizations of model (6), and average the smoothed periodograms to represent the ground truth in Figure 1(f) and (g). Notably, while the PG of model (6) is a constant, the EPs detect the low-frequency feature at both the lower and higher expectiles.

The second example is based on electroencephalogram (EEG) data collected from an epilepsy patient during a seizure interval[1]. Figure 2(a) presents the EEG segments along with their corresponding sample expectiles. The PG in Figure 2(b) successfully detects both low-frequency and high-frequency components, corresponding to the six main spikes and the associated bursts, respectively. The EPs in Figure 2(b) and (c) provide additional information, demonstrating that the high-frequency patterns appear predominantly at higher expectile levels. This observation is consistent with the EEG data, where the bursts primarily occur at the top of the main spikes. Furthermore, the intensity of the bursts increases with higher expectiles, particularly at frequencies $\omega \in [0.02, 0.03] \times 2\pi$, as shown in Figure 2(d).

3 Asymptotic Analysis

3.1 Large Sample Theory

It is well-established that a smoothed PG constitutes an estimate of the power spectrum of the underlying random process that generates the time series. In this section, we conduct a parallel investigation of the EP through large-sample asymptotic analysis. Our results reveal that the EP is fundamentally connected to a novel spectrum-type function we term the expectile spectrum.

Let $\{y_t\}$ be a stationary process with cumulative distribution function $F(\cdot)$ and finite second moments. Define

$$\boldsymbol{\zeta}_n(\alpha) := n^{-1/2} \sum_{t=1}^{n} \dot{\rho}_\alpha(y_t - \mu(\alpha))\mathbf{x}_{nt}, \tag{7}$$

where the regressor $\mathbf{x}_{nt}$ is defined in (4). Assume that the following conditions are satisfied for fixed α:

[1]The data was collected by South China Normal University School of Psychology and is authorized for use in this paper.

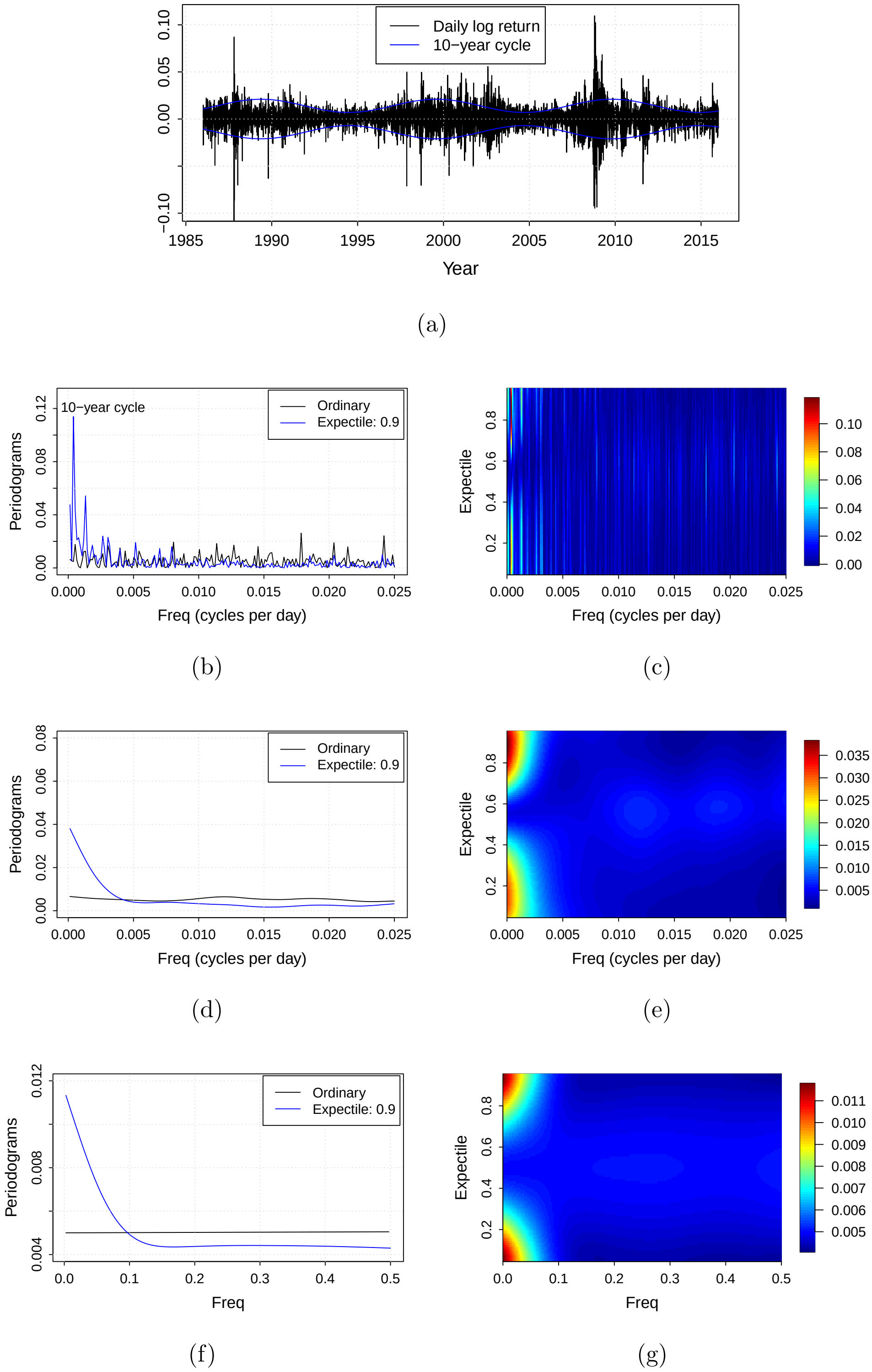


Figure 1: (a) Daily log returns of the S&P 500 Index data; (b) the PG and the EP at expectile $\alpha = 0.9$; (c) the EP at expectiles $\{0.05, 0.06, ..., 0.95\}$; (d) and (e) the smoothed versions of (b) and (c), respectively; (f) the averaged PG and the averaged EP at expectile $\alpha = 0.9$ of model (6); and (g) the averaged EP of model (6) at expectiles $\{0.05, 0.06, ..., 0.95\}$. (f) and (g) are ensemble means of 5,000 realizations.

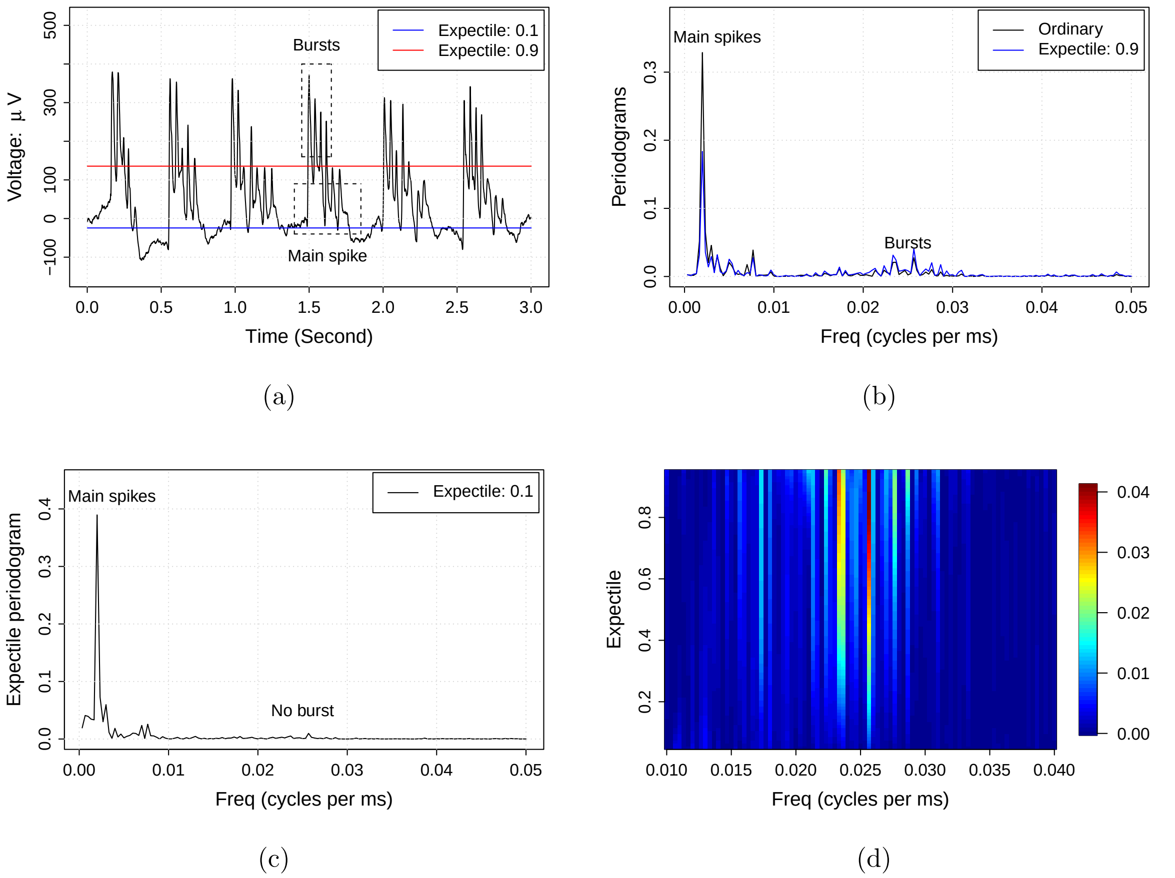


Figure 2: (a) The EEG data over a 3-second interval, along with its sample expectiles 0.1 and 0.9; (b) the PG and the EP at $\alpha = 0.9$; (c) the EP at $\alpha = 0.1$; and (d) the EP at expectiles $\{0.05, 0.06, ..., 0.95\}$. The time resolution of the EEG data is 1000 Hz.

C1. The cumulative distribution function $F(\cdot)$ is Lipschitz continuous, i.e., there exists a constant $K > 0$ such that $|F(y) - F(y')| \le K|y - y'|$ for all $y, y' \in \mathbb{R}$.

C2. The process $\{y_t\}$ has the strong mixing property with mixing numbers a_τ ($\tau = 1, 2, \ldots$) satisfying $a_\tau \to 0$ as $\tau \to \infty$.

C3. [2] A central limit theorem is valid for $\{\boldsymbol{\zeta}_n(\alpha)\}$, i.e., $\boldsymbol{\zeta}_n(\alpha) \xrightarrow{D} N(\mathbf{0}, \mathbf{V}(\alpha))$ as $n \to \infty$, where $\mathbf{V}(\alpha)$ is the asymptotic covariance matrix of $\boldsymbol{\zeta}_n(\alpha)$.

The sequence $\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\}$ is stationary under assumption (C2). Therefore, we define the autocovariance function (ACF) of $\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\}$ as:

$$\mathrm{Cov}\{\dot{\rho}_\alpha(y_t - \mu(\alpha)), \dot{\rho}_\alpha(y_s - \mu(\alpha))\} = \gamma(t - s, \alpha).$$

Assume that $\sum_\tau |\gamma(\tau, \alpha)| < \infty$, define

$$g(0, \alpha) = \eta^2(\alpha) h(0, \alpha) \text{ and } g(\omega, \alpha) = \eta^2(\alpha) h(\omega, \alpha), \tag{8}$$

[2]Condition C3 can be substituted by some other conditions, see Section 8.4 in **Appendix** for details.

where $h(0,\alpha) := \sum_\tau \gamma(\tau,\alpha)$, and

$$h(\omega,\alpha) := \sum_{\tau=-\infty}^{\infty} \gamma(\tau,\alpha)\exp(-i\omega\tau) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau,\alpha)\cos(\omega\tau). \tag{9}$$

The second expression in (9) is due to the symmetry $\gamma(\tau,\alpha) = \gamma(-\tau,\alpha)$. The scaling factor $\eta^2(\alpha)$ is obtained by the normal equation (2). Because

$$\dot{\rho}_\alpha(u) := \begin{cases} 2\alpha u, & \text{if } u \geq 0, \\ 2(1-\alpha)u & \text{otherwise,} \end{cases}$$

the normal equation (2) is equivalent to

$$(1-\alpha)\int_{-\infty}^{\mu(\alpha)} (y-\mu(\alpha))\, dF(y) + \alpha \int_{\mu(\alpha)}^{\infty} (y-\mu(\alpha))\, dF(y) = 0,$$

or

$$\mu(\alpha)\left\{(1-\alpha)F(\mu(\alpha)) + \alpha(1-F(\mu(\alpha)))\right\} = (1-\alpha)\int_{-\infty}^{\mu(\alpha)} y\, dF(y) + \alpha\int_{\mu(\alpha)}^{\infty} y\, dF(y).$$

In the special case of $\alpha = 0.5$, the normal equation reduces to

$$\int (y-\mu(0.5))dF(y) = 0,$$

which yields $\mu(0.5) = \mathbb{E}\{y_t\}$. The quantity

$$\alpha(1-F(\mu(\alpha))) + (1-\alpha)F(\mu(\alpha)) = \alpha + (1-2\alpha)F(\mu(\alpha))$$

always lies between α and $1-\alpha$. Therefore,

$$\eta(\alpha) := \left[\frac{1}{2}\left\{\alpha(1-F(\mu(\alpha))) + (1-\alpha)F(\mu(\alpha))\right\}\right]^{-1} \tag{10}$$

is a finite positive number for any given $\alpha \in (0,1)$.

Equipped with these concepts, we establish the following theorem regarding the asymptotic normality of the trigonometric ER solution $\widehat{\boldsymbol{\beta}}_n(\omega_\nu,\alpha)$ and the expectile periodogram $EP_n(\omega_\nu,\alpha)$ in (5). A proof is given in the Appendix.

Theorem 1 (Trigonometric Expectile Regression and Expectile Periodogram). Assume that the ACF $\gamma(\tau,\alpha)$ of $\{\dot{\rho}_\alpha(y_t-\mu(\alpha))\}$ is absolutely summable with $h(0,\alpha) > 0$ and $h(\omega,\alpha) > 0$. If $\omega_\nu \to \omega \in (0,\pi)$ as $n \to \infty$. Then, under (C1)-(C3), we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n(\omega_\nu,\alpha) - \hat{\boldsymbol{\beta}}_0(\alpha)) \xrightarrow{D} N(\mathbf{0}, \mathrm{diag}\{g(0,\alpha), 2g(\omega,\alpha), 2g(\omega,\alpha)\}), \tag{11}$$

and

$$EP_n(\omega_\nu, \alpha) \xrightarrow{D} g(\omega, \alpha)(1/2)\chi_2^2, \tag{12}$$

where $g(0, \alpha)$ and $g(\omega, \alpha)$ are defined in (8), and χ_2^2 denotes a random variable having the χ^2 distribution with 2 degrees of freedom.

Consider the case for multiple frequencies, for fixed $q > 1$, consider $\{\mathbf{x}_{jnt} : t = 1, \ldots, n, j = 1, \ldots q\}$ with $\mathbf{x}_{jnt} := [1, \cos(\omega_{\nu_j} t), \sin(\omega_{\nu_j} t)]^T$. The quantity $\boldsymbol{\zeta}_n(\alpha)$ in (7) generalizes to

$$\boldsymbol{\zeta}_{jn}(\alpha) := n^{-1/2} \sum_{t=1}^{n} \dot{\rho}_\alpha(y_t - \mu(\alpha))\mathbf{x}_{jnt}. \tag{13}$$

With the above definitions, we extent **Theorem 1** to the following result.

Theorem 2 (Expectile Periodogram). For fixed $q > 1$ and $0 < \lambda_1 < \cdots < \lambda_q < \pi$, let $\omega_{\nu_1}, \ldots, \omega_{\nu_q}$ be Fourier frequencies satisfying $\omega_{\nu_j} \to \lambda_j$ as $n \to \infty$ for $j = 1, \ldots, q$. If the ACF $\gamma(\tau, \alpha)$ of $\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\}$ is absolutely summable with $h(\lambda_j, \alpha) > 0$ $(j = 1, \ldots, q)$, and if (C1), (C2) and

C3*. A central limit theorem is valid for $\text{vec}[\boldsymbol{\zeta}_{jn}(\alpha)]_{j=1}^q$ for all j, i.e., $\text{vec}[\boldsymbol{\zeta}_{jn}(\alpha)]_{j=1}^q \xrightarrow{D} N(\mathbf{0}, [\mathbf{V}_{jk}(\alpha)]_{j,k=1}^q$ as $n \to \infty$, where $\mathbf{V}_{jk}(\alpha)$ is the asymptotic covariance matrix of $\boldsymbol{\zeta}_{jn}(\alpha)$ and $\boldsymbol{\zeta}_{kn}(\alpha)$.

are satisfied, then

$$\{EP_n(\omega_{\nu_1}, \alpha), \ldots, EP_n(\omega_{\nu_q}, \alpha)\} \xrightarrow{D} \{g(\lambda_1, \alpha)(1/2)\chi_{2,1}^2, \ldots, g(\lambda_q, \alpha)(1/2)\chi_{2,q}^2\}, \tag{14}$$

where $\chi_{2,1}^2, \ldots, \chi_{2,q}^2$ are i.i.d. random variables each having a χ^2 distribution with 2 degrees of freedom.

Theorems 1 and 2 show that the EP exhibits scaled chi-square distributions with two degrees of freedom, which is similar to the behavior of the PG (Brockwell and Davis, 2013).

3.2 Asymmetrically-Scaled Process and Expectile Spectrum

The definition of $h(\omega, \alpha)$ in (9) can be extended to include all $\omega \in [0, 2\pi)$. As such a function of ω, it is nothing but the ordinary spectrum of the stationary process $\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\}$. This process is just a asymmetrically-scaled version of the process $\{y_t - \mu(\alpha)\}$

around the α-expectile $\mu(\alpha)$. We term $\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\}$ as the asymmetrically-scaled expectile crossing process (ASECP) of y_t, which corresponds to the level-crossing process (LCP) $\{I(y_t > \alpha)\}$ that associate with the quantile spectrum.

Similarly, as a scaled version of $h(\omega, \alpha)$, the definition of $g(\omega, \alpha)$ can also be extended to include all $\omega \in [0, 2\pi)$. We define this function the α-expectile spectrum of $\{y_t\}$. According to (10), the scaling function $\eta^2(\alpha)$ depends solely on the marginal distribution $F(\cdot)$.

Note that the ACF of $\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\}$ can be written as:

$$
\begin{aligned}
\gamma(\tau, \alpha) &= \mathrm{E}\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\, \dot{\rho}_\alpha(y_{t-\tau} - \mu(\alpha))\} \\
&= \alpha^2 \mathrm{E}\{(y_t - \mu(\alpha))(y_{t-\tau} - \mu(\alpha)) I(y_t > \mu(\alpha), y_{t-\tau} > \mu(\alpha))\} \\
&\quad + (1-\alpha)^2 \mathrm{E}\{(y_t - \mu(\alpha))(y_{t-\tau} - \mu(\alpha)) I(y_t \le \mu(\alpha), y_{t-\tau} \le \mu(\alpha))\} \\
&\quad + 2\alpha(1-\alpha) \mathrm{E}\{(y_t - \mu(\alpha))(y_{t-\tau} - \mu(\alpha)) I(y_t > \mu(\alpha), y_{t-\tau} \le \mu(\alpha))\}.
\end{aligned}
$$

In the special case of $\alpha = 0.5$, it reduces to the ordinary ACF of $\{y_t\}$.

Figure 3(a) and (b) illustrate an realization of model (16) with $\omega_{\nu_c} = 0.25 \times 2\pi$, along with the corresponding 0.8 expectile and quantile, respectively. As shown in Figure 3(c) and (d), the ASECP contains more information, as the LCP losses information when transforming a real-valued time series into a binary-valued level-crossing process. This is further supported by Figure 3(c) and (d), where the averaged EP is smoother than QP, indicating a faster convergence rate and higher statistical efficiency for the EP.

4 Numerical Results

In this section, we present simulations to demonstrate the efficiency of the EP in detecting hidden periodicities in time series.

4.1 Hidden Periodicities Detection

We consider the following model:

$$
y_t = a_t x_t, \tag{15}
$$

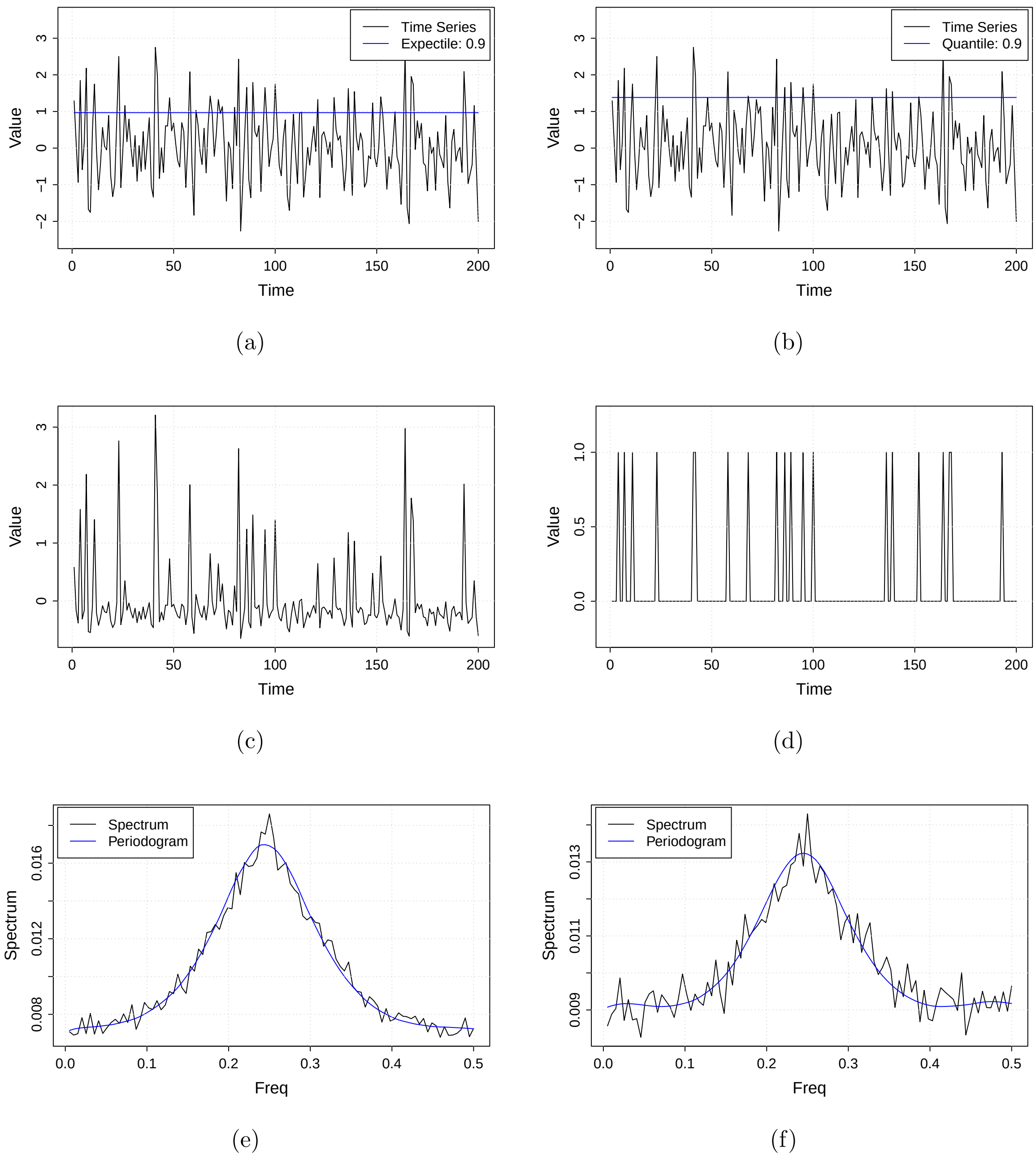


Figure 3: Left panel: expectile; right panel: quantile. Top row: time series and their corresponding 0.8 expectile and quantile; middle row: the ASECP and LCP of the time series; bottom row: the spectra and the averaged periodograms of 500 realizations. The spectra are computed by averaging 5,000 smoothed periodograms.

where

$$a_t = b_0 + b_1 \cos(\omega_{\nu_0} t) + b_2 \sin(\omega_{\nu_1} t),$$

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

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t, \tag{16}$$

where $\phi_1 = 2r\cos(\omega_{\nu_c}), \phi_2 = -r^2$ ($r = 0.6$) and $\{\epsilon_t\}$ is the noise. Additionally, we set $\omega_{\nu_0} = 0.1 \times 2\pi, \omega_{\nu_1} = 0.12 \times 2\pi$. This setup aims to evaluate the effectiveness of different types of periodograms in detecting multiple closely spaced periodicities.

We first present the periodograms of model (16) with $\omega_{\nu_c} = 2\pi \times 0.25$ and $2\pi \times 0.3$, representing cases without hidden periodicities. As shown in Figure 4(a) and 4(b), the PGs, the EPs and the Laplace periodograms, exhibit similar bell-shaped patterns, with spectral peaks around ω_c. These results are consistent with the spectral properties of AR(2) process. Specifically, the characteristic polynomial of model (16) is $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$, where the roots z_1 and z_2 are complex conjugates. The magnitude $|z_1| = |z_2| = 1/r > 1$ ensures causality. The AR coefficients ϕ_1 and ϕ_2 determine the location and narrowness of the spectral peak. Specifically, the peak frequency is located at $\omega_c/2\pi$ and as $r \to 1^-$, the peak becomes narrower (Shumway and Stoffer, 2016). Figure 4(c) and (d) show the EPs at expectiles $\{0.05, 0.06, ..., 0.95\}$, which share similar features to the quantile spectrum in Chen et al. (2021b). Specifically, all periodograms shown are ensemble means of 5,000 smoothed periodograms.

Figure 5 illustrates the ability of the EP in detecting hidden periodicities. We present the ensemble mean of 5,000 realizations of the PGs and EPs of model (15). The EPs detect the hidden periodicities, manifesting as large spikes at ω_{ν_0} and ω_{ν_1}, whereas the PGs do not exhibit this capability. It is important to note that spectral leaks, which have been observed for the Laplace periodogram, can also occur in the EP. Therefore, small spikes may take place at some other frequencies, as indicated by Theorem 2 in Li (2012b). To address the issue of spectral leakage, incorporating an l_1 regularization into the loss function is beneficial (Meziani et al., 2020).

The EPs in Figure 1(g) and Figure 4(c), (d) are symmetric across the expectile level. To demonstrate the capability of the EP in handling more complex spectral features, we construct y_t, whose EP is intentionally made asymmetric across the expectile level. y_t is

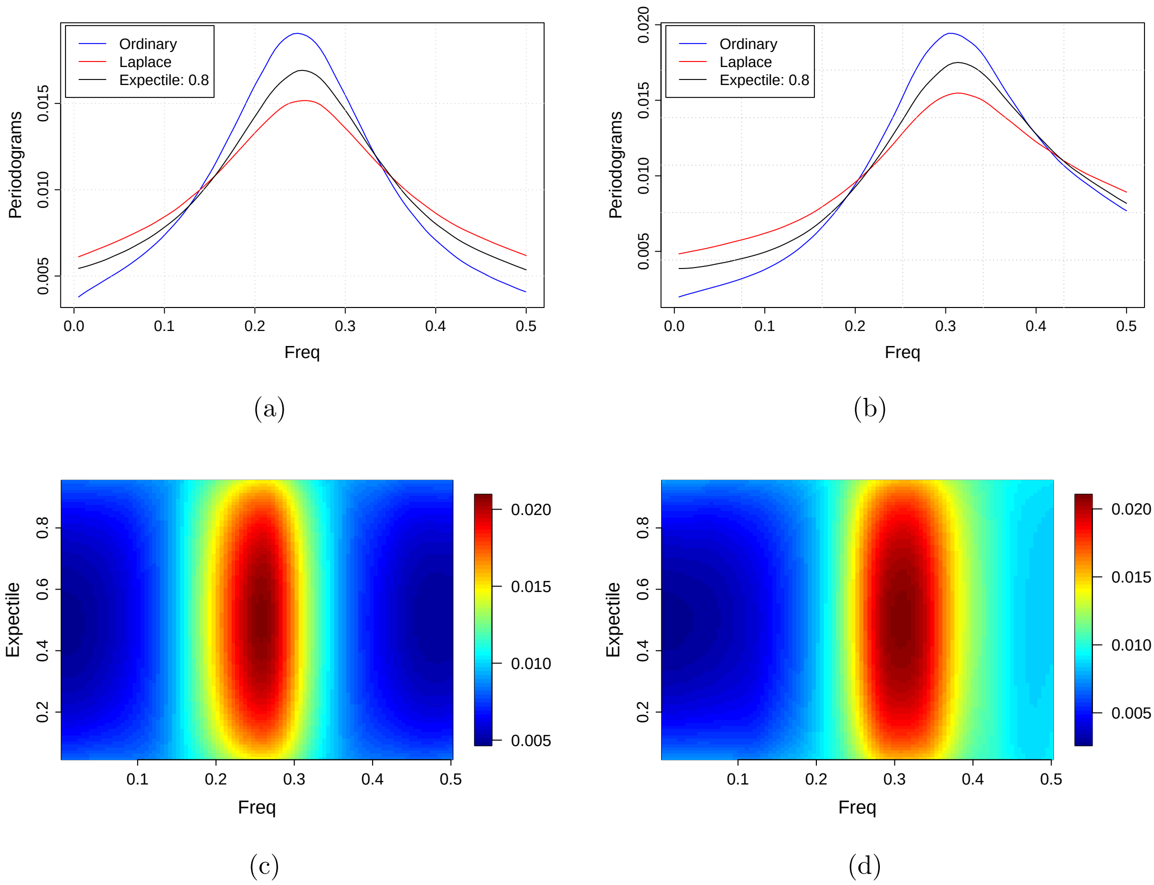


Figure 4: Ensemble means of three types of smoothed periodograms of time series defined by model (16) with standard Gaussian white noise. (a) Periodograms with $\omega_{\nu_c} = 2\pi \times 0.25$, (b) periodograms with $\omega_{\nu_c} = 2\pi \times 0.3$, (c) and (d) the EPs at expectiles $\{0.05, 0.06, ..., 0.95\}$ for $\omega_{\nu_c} = 2\pi \times 0.25$ and 0.3, respectively. The number of realizations is 5,000 and the sample size $n = 200$.

defined as a nonlinear mixture of three components:

$$z_t = W_1(x_{t,1})\, x_{t,1} + \{1 - W_1(x_{t,1})\}\, x_{t,2},$$

$$y_t = W_2(z_t)\, z_t + \{1 - W_2(z_t)\}\, x_{t,3}. \tag{17}$$

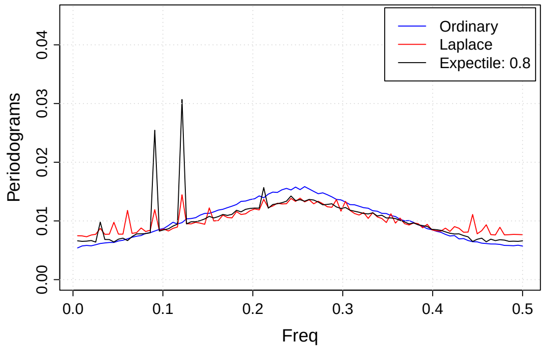
The components $\{x_{t,1}\}$, $\{x_{t,2}\}$, and $\{x_{t,3}\}$ are independent Gaussian AR(1) processes satisfying

$$x_{t,1} = 0.8x_{t-1,1} + w_{t,1},$$

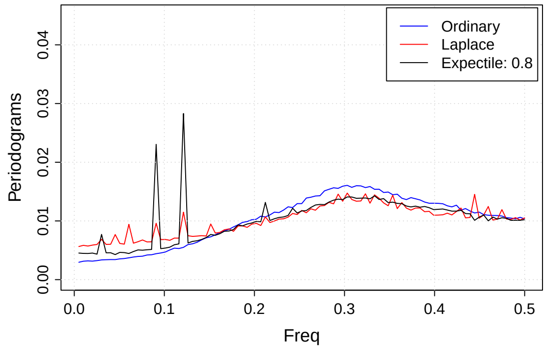
$$x_{t,2} = -0.75x_{t-1,2} + w_{t,2},$$

$$x_{t,3} = -0.81x_{t-2,3} + w_{t,3},$$

where $w_{t,1}, w_{t,2}, w_{t,3}$ are standard Gaussian white noise. From the perspective of traditional spectral analysis, the series $\{x_{t,1}\}$ has a lowpass spectrum, $\{x_{t,2}\}$ has a highpass



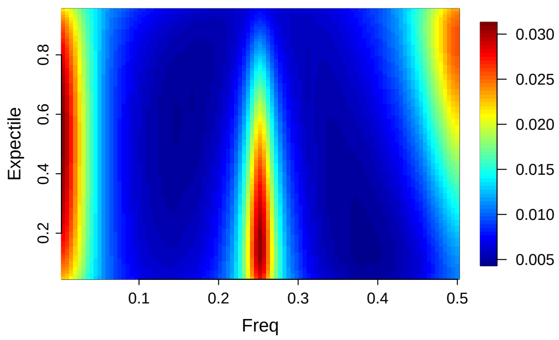
(a)



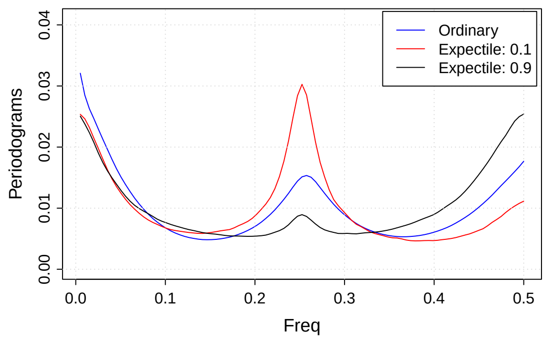
(b)

Figure 5: Means of the three types of periodograms of time series defined by model (15). (a) Represents $\omega_{\nu_c} = 0.25$ and (b) represents $\omega_{\nu_c} = 0.3$, respectively. The number of realizations is $5,000$ and $n = 200$.

spectrum, and $\{x_{t,3}\}$ has a bandpass spectrum around frequency $1/4$. The mixing function $W_1(x)$ is equal to 0.9 for $x < -0.8$, 0.2 for $x > 0.8$, and linear transition for x in between. The mixing function $W_2(x)$ is similarly defined except that it equals 0.5 for $x < -0.4$ and 1 for $x > 0$. Figure 6(a) shows the EPs of model (17), where the EP is asymmetric across the expectile levels. Figure 6(b) compares the PGs and the EPs at $\alpha = 0.1$ and 0.9, illustrating that the PG is restricted to capturing spectral features near the central expectiles. In contrast, the EP offers a broader perspective, effectively analyzing the time series across the entire range of $\alpha \in (0, 1)$.



(a)



(b)

Figure 6: The periodograms of the mixture model (17). (a) The EP with asymmetric pattern across the expectile level, and (b) the EPs at extreme expectiles ($\alpha = 0.1$ and 0.9), as well as the PG. The number of realizations is 5,000 and the sample size $n = 200$.

4.2 Fisher's Test

One commonly used hypothesis test to detect periodicities for the PG is Fisher's test (Brockwell and Davis, 1991). For frequencies $\{\omega_{\nu_1}, \omega_{\nu_2}, ..., \omega_{\nu_q}\}$, the test statistic is defined as

$$g = \frac{\max_{1 \le k \le q}\{I_n(\omega_{\nu_k})\}}{\sum_{k=1}^{q} I_n(\omega_{\nu_k})}.$$

According to Fisher's test, a sufficiently large value of g indicates the presence of a hidden periodicity. The null hypothesis is that the time series is Gaussian white noise, while the alternative hypothesis is that the time series contains a deterministic periodic component of unspecified frequency.

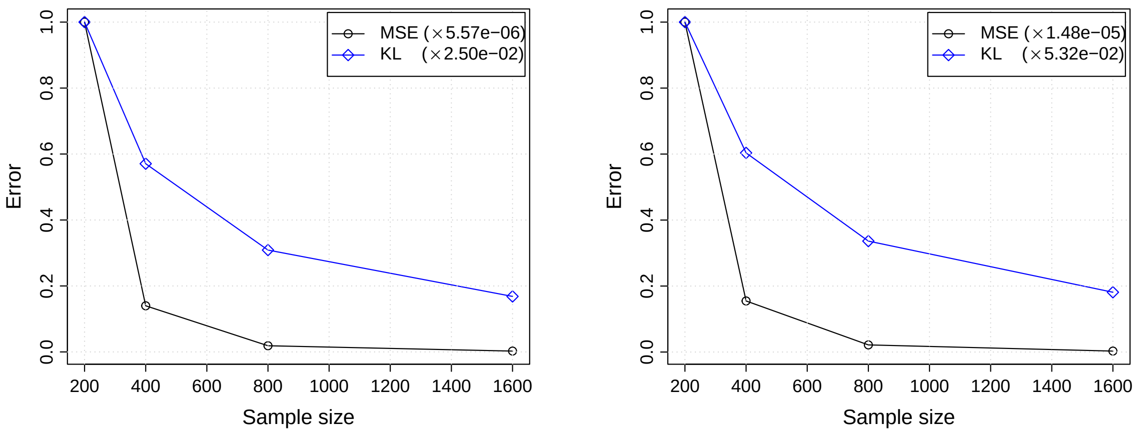
We apply Fisher's test to the EP by replacing $I_n(\omega)$ with $EP_n(\omega, \alpha)$ in the test statistic. The probabilities of detection are obtained by 1,000 Monte Carlo simulation runs for time series defined by model (15), with $\omega_{\nu_0} = 0.1 \times 2\pi, \omega_{\nu_c} = 0.3 \times 2\pi$, and $b_2 = 0$. As can be seen in Table 1, both the EPs and QPs outperform the PG. At a significance level of 0.05, the EP ($\alpha = 0.9$) achieves a detection rate of 84.26%, whereas the PG only achieves only 29.78%. Detection rates are sensitive to the expectile level, showing a clear trend in this experiment. As the expectile or quantile approaches 1, the detection rate of the EP increases and eventually surpasses that of the QP.

Table 1: Fisher's test of different types of periodograms.

Significa-	EP			QP			PG
nce level	α=0.85	α=0.9	α=0.95	α=0.85	α=0.9	α=0.95	
0.01	0.4048	0.5608	0.5850	0.6898	0.6328	0.4428	0.1224
0.05	0.7158	0.8426	0.8510	0.8720	0.8260	0.6646	0.2978
0.10	0.8308	0.9262	0.9306	0.9278	0.8952	0.7678	0.4258

4.3 Smoothed EP

It is well known that the power spectrum can be estimated consistently by a properly smoothed PG (Brockwell and Davis, 1991). Theorems 1 and 2 suggests that the expectile spectrum, defined by (8), can be estimated nonparametrically by smoothing the EP ordinates at the Fourier frequencies in the same way. The simulation studies confirm that the estimation accuracy decreases with the increase of n. We measure the distance between the smoothed EP and the expectile spectrum (the average of 5,000 smoothed EPs) using the mean squared error (MSE) and Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951). Models (16) and (17) at $\alpha = 0.9$ are used for illustration. We smooth the EP at a certain expectile using splines (function smooth.splines in R, with tuning parameter cv = TRUE). Figure 7 depicts the MSE and KL divergence as a function of n, with n taking values of 200, 400, 800, and 1,600. To highlight the decreasing trend as n increases, the values are scaled by setting the maximum to 1.



(a) (b)

Figure 7: MSE and KL divergence of smoothed periodogram. (a) AR(2) model (16); (b) mixture model (17). The results are based on 5,000 simulation runs.

5 Earthquake Data Classification

In this section, we apply the EP to an earthquake classification problem. Section 5.1 introduces the earthquake data and Section 5.2 describes the deep learning model and presents the classification results.

5.1 Data Description

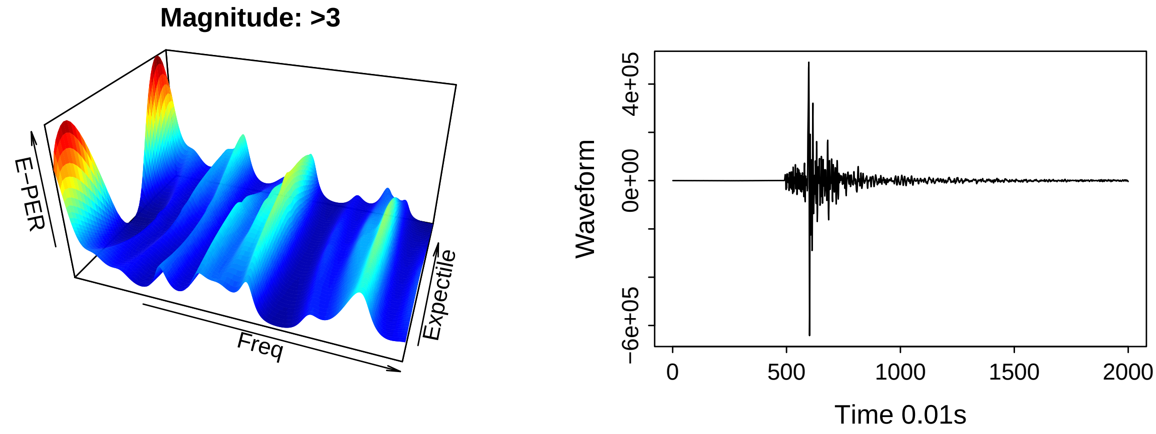
The earthquake waveform data with a sampling rate of 100 Hz, was collected in February 2014 in Oklahoma State. This data is available at `https://www.iris.edu/hq/` and http://www.ou.edu/ogs.html. Details about the catalog data are provided in Benz et al. (2015), including labels for earthquake magnitudes and occurrence times. We extract 2,000 non-overlapping segments, each being a time series with $n = 2,000$, corresponding to 20 seconds of data. The choice of longer time series ensures that the segments contain complete earthquake events. Among these segments, 1,000 of them contain an earthquake with magnitudes greater than 0.25, while the remaining 1,000 segments contain no earthquakes. We smooth the EPs of the 2,000 segments using the semi-parametric method proposed in Chen et al. (2021b), ensuring smoothness across both the expectile and frequency dimensions. In this experiment, we use the lower half of the frequencies ($\omega \leq 0.25 \times 2\pi$) and 46 expectiles $\{0.05, 0.07, ..., 0.93, 0.95\}$. Since we focus on serial dependence and normalize the EPs, amplitude considerations are excluded, making the classification more challenging. Additionally, we incorporate two competitive periodograms: the smoothed PG and QP.

We show three representative segments along with their corresponding smoothed EPs in Figure 8. Specifically, Figure 8(a) contains a large earthquake with a magnitude larger than 3, Figure 8(b) contains a somehow small earthquake with a magnitude less than 1, and Figure 8 (c) contains no earthquake. Based on the three segments, we observe the following features:

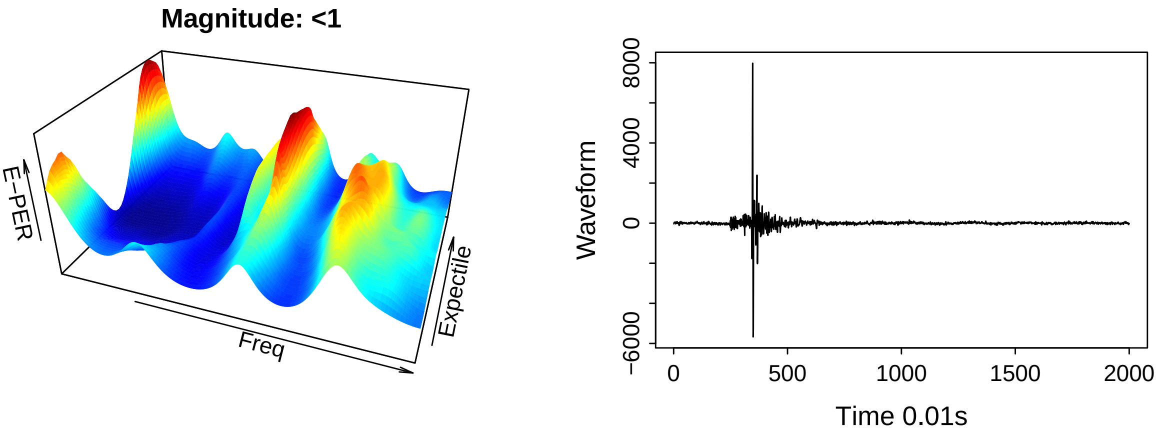
- The smoothed EP for the segment with a large earthquake exhibits spectral peaks at the low-frequency band at both higher and lower expectiles.
- The smoothed EP for the segment with a small earthquake exhibits spectral peaks at both the low-frequency band (at lower and higher expectiles) and the high-frequency band (at middle expectiles).
- The smoothed EP for the segment with no earthquake exhibits peaks only in the high-frequency bands.

5.2 Classification using Deep Learning Model

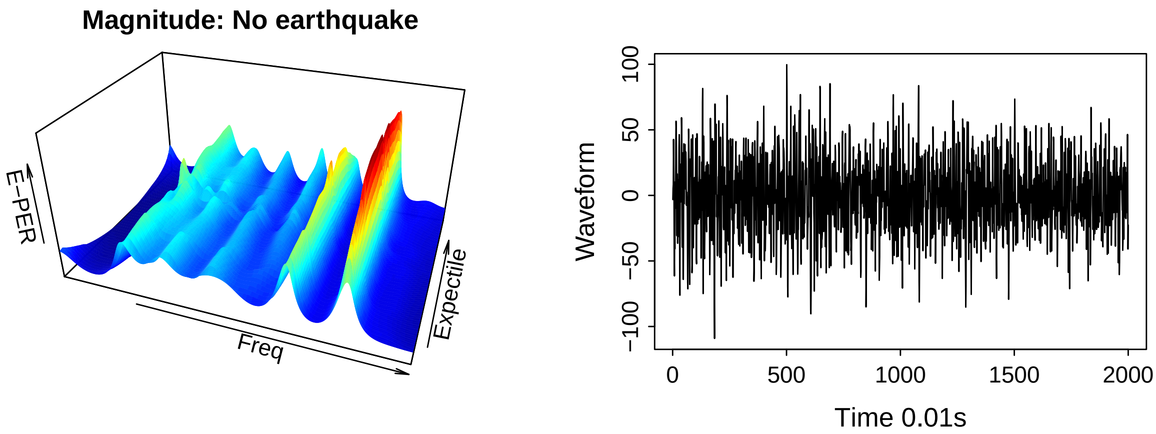
In this section, we use the three types of smoothed periodograms as features to classify the segments into those that contain earthquakes and those that do not. We randomly



(a)



(b)



(c)

Figure 8: Three segments and the corresponding smoothed EPs. (a) The segment with an earthquake with a magnitude > 3, (b) the segment with an earthquake with a magnitude <1, and (c) the segment with no earthquake. $n = 2000$, $\omega \leq 0.25 \times 2\pi$, and $\alpha = 0.05, 0.07, ..., 0.95$.

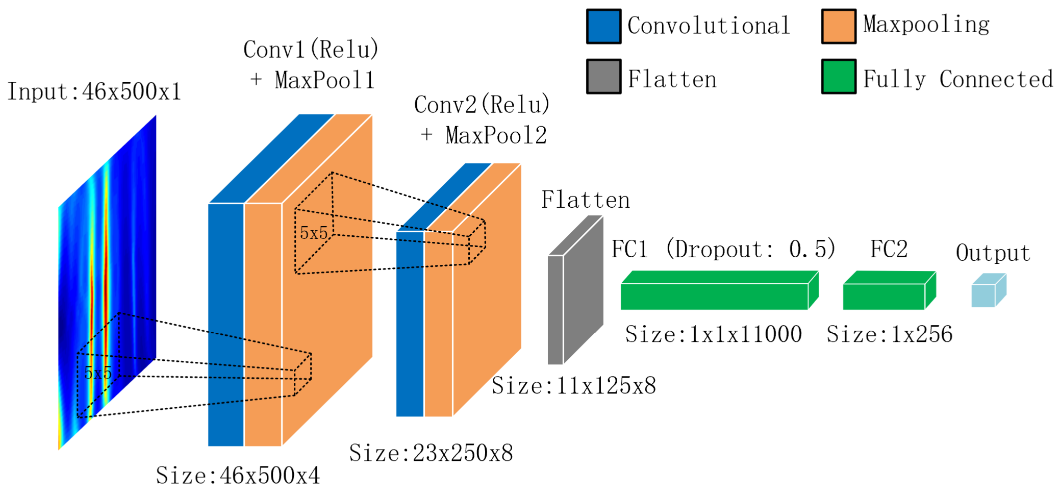


Figure 9: Structure of the deep learning model.

split the segments into training and testing sets, comprising 1,600 and 400 samples, respectively. To classify the EPs and QPs, we employ a model with two convolutional layers for feature extraction, each followed by a max-pooling layer. The second pooling layer connects to a fully connected (FC) layer after flattening the output. A dropout layer with a rate of 50% is applied to the FC layer, leading to the output layer. The total number of trainable parameters is 2,817,682, and the learning rate is set to be 10^{-4} with a reduction rate of 0.5 every 20 epochs. Additional details about the model are presented at `https://github.com/tianbochen1`. The model structure is illustrated in Figure 9. To classify the PGs, which are one-dimensional with respect to ω, we adapt the model to use different input dimensions (1×500 instead of 46×500) and kernel size (1×5 instead of 5×5).

We conduct the training ten times, each with a randomly constructed training–testing split and randomly initialized weights. Over 80% of the training procedures converge within 30 epochs. The testing accuracies for the three types of periodograms are as follows:

- EP: {0.9900, 1.0000, 0.9925, 1.0000, 0.9975, 0.9950, 0.9925, 0.9925, 0.9950,0.9925}.
- QP: {0.9825, 0.9900, 0.9925, 0.9925, 0.9825, 0.9950, 0.9925, 0.9900, 0.9875, 0.9825}.
- PG: {0.9875, 0.9725, 0.9825, 0.9825, 0.9975, 0.9850, 0.9850, 0.9825, 0.9775, 0.9900}.

The averaged confusion matrices are shown in Table 2, in which true positive (TP) indicates that a segment with an earthquake is correctly classified as containing an earthquake, true negative (TN) indicates that the segment without an earthquake is correctly classified as not containing an earthquake; false positive (FP) indicates that a segment without an earthquake is incorrectly classified as containing one; and false negative (FN) indicates that a segment with an earthquake is incorrectly classified as not containing one (where P denotes positive, N denotes negative, T denotes true, and F denotes false).

Table 3 shows three classification metrics: accuracy, precision ($\frac{TP}{TP+FP}$), and recall ($\frac{TP}{TP+FN}$). The optimal value in each row for the three types of periodograms is highlighted in bold. Additionally, we present the time required for estimation and training (per epoch) to compare the computational complexity of the expectile and QPs. All computations were performed on a PC equipped with an Intel Core i9-13900KF CPU @5.2GHz, 64 GB of RAM, and an Nvidia RTX 4090 GPU.

Table 2: The averaged confusion matrices of the classification.

(a) EP

	P	N
T	199.3	198.6
F	0.7	1.4

(b) QP

	P	N
T	196.2	199.3
F	3.8	0.7

(c) PG

	P	N
T	196.2	197.5
F	3.8	2.5

Table 3: The classification results.

Metrics		EP	QP	PG
Accuracy	Averaged	**0.9948**	0.9880	0.9858
	[min, max]	[0.9925, 1.0000]	[0.9725, 0.9950]	[0.9750, 0.9925]
Precision	Averaged	**0.9965**	0.9840	0.9843
	[min, max]	[0.9896, 1.0000]	[0.9559, 0.9952]	[0.9609, 0.9902]
Recall	Averaged	**0.9931**	0.9921	0.9877
	[min, max]	[0.9858, 1.0000]	[0.9794, 1.0000]	[0.9653, 1.0000]
Time (s)	Estimation	12.6684	19.4528	–
	Training	0.3739	0.3731	–

From the results, we can see that:

- The classification based on the EP has higher testing accuracy, precision, and recall rate than both quantile and PGs. This indicates that the EPs are suitable features for time series classification.

- One misclassification case in FN using the EPs is shown in Figure 10. Since the magnitude is too small (< 0.1), the power at low frequencies is not as large as the power at high frequencies, which causes the misclassification.

- The EP incurs a lower estimation complexity than the QP. However, its computational cost is higher than that of the PG, as the dimension is multiplied by the number of expectile levels.

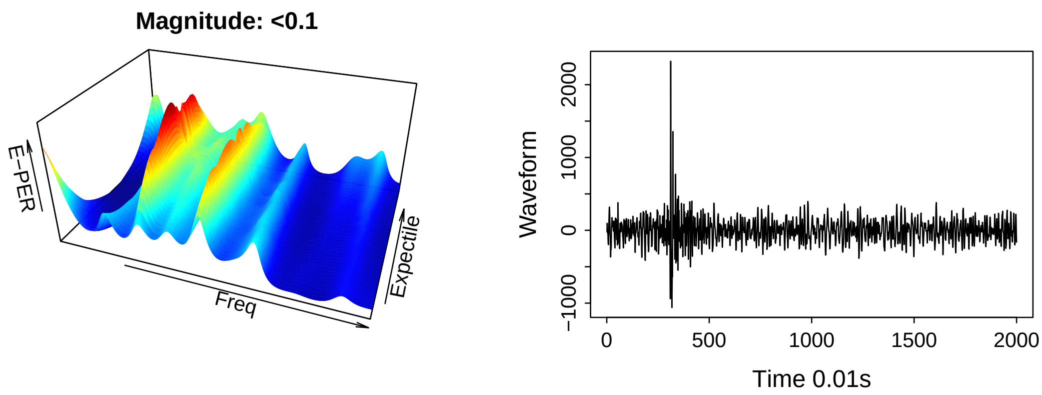


Figure 10: A misclassification case of FN.

6 Conclusion

We have proposed the EP as a counterpart to the PG and investigated its potential as a non-parametric tool for time series analysis. We also establish the asymptotic theory and investigate the relationship between the EP and what we defined as expectile spectrum. The EP offers more information than the PG by examining the serial dependence at different expectile levels. We conduct real-world examples and simulation studies to highlight their proficiency in detecting hidden periodicities within time series. In the earthquake data classification task, we leverage the inherent two-dimensional characteristics of the EP using a deep learning model, which is a powerful technique in image classification.

Despite its computational efficiency relative to the QP, the EP demands higher computational resources than the PG. This computational burden arises from the fact that the dimensionality is multiplied by the number of expectiles used. One solution to reduce this computational cost is to use fewer expectiles. In this paper, we choose a large number of expectiles uniformly across $(0, 1)$. Researchers may choose to focus exclusively on a subset of expectiles with sufficient discriminative power (e.g., high or low expectiles). Furthermore, we utilize multi-thread parallelization to speed up computation using the package foreach, doParallel in R.

In closing, the further developments are needed. First, it is worthwhile to derive the joint asymptotic distribution of the EP at multiple quantile levels to better understand the behavior of the EP for expectile-frequency analysis. Second, investigating the limiting properties of the EP at extremely high expectiles maybe helpful. At last, developing smoothing algorithms for the EP would enhance its practical utility.

7 Supplemental Material

The code for estimating the EP and reproducing the results in Section 4 and Section 5 is accessible at: `https://github.com/tianbochen1/Expectile-Periodogram`.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant No.12301326 and No.12301171; Anhui Provincial Natural Science Foundation under Grant No.2308085QA05 and No.2308085QA04; the University Natural Science Research Project of Anhui Province under Grant No.2023AH050099.

References

Alvarado, R., Tillaguango, B., Dagar, V., Ahmad, M., Işık, C., Méndez, P., and Toledo, E. (2021). Ecological footprint, economic complexity and natural resources rents in latin america: empirical evidence using quantile regressions. *Journal of Cleaner Production*, 318:128585.

Baud, M. O., Kleen, J. K., Mirro, E. A., Andrechak, J. C., King-Stephens, D., Chang,

E. F., and Rao, V. R. (2018). Multi-day rhythms modulate seizure risk in epilepsy. *Nature communications*, 9(1):88.

Benz, H. M., McMahon, N. D., Aster, R. C., McNamara, D. E., and Harris, D. B. (2015). Hundreds of earthquakes per day: The 2014 guthrie, oklahoma, earthquake sequence. *Seismological Research Letters*, 86(5):1318–1325.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Birr, S., Volgushev, S., Kley, T., Dette, H., and Hallin, M. (2017). Quantile spectral analysis for locally stationary time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1619–1643.

Bloomfield, P. (2004). *Fourier analysis of time series: an introduction*. John Wiley & Sons.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer.

Brockwell, P. J. and Davis, R. A. (2013). *Time series: theory and methods*. Springer Science & Business Media.

Cai, Z. and Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, 167(2):413–425.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103(484):1595–1608.

Caiado, J., Crato, N., and Peña, D. (2006). A periodogram-based metric for time series classification. *Computational Statistics & Data Analysis*, 50(10):2668–2684.

Chen, T., Sun, Y., Euan, C., and Ombao, H. (2021a). Clustering brain signals: A robust approach using functional data ranking. *Journal of Classification*, 38:425–442.

Chen, T., Sun, Y., and Li, T.-H. (2021b). A semi-parametric estimation method for the quantile spectrum with an application to earthquake classification using convolutional neural network. *Computational Statistics & Data Analysis*, 154:107069.

Dette, H., Hallin, M., Kley, T., Volgushev, S., et al. (2015). Of copulas, quantiles, ranks and spectra: An l_1 -approach to spectral analysis. *Bernoulli*, 21(2):781–831.

Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, pages 93–125.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4):987–1007.

Euán, C., Ombao, H., and Ortega, J. (2018). The hierarchical spectral merger algorithm: a new time series clustering procedure. *Journal of Classification*, 35:71–99.

Garcia, J., Hernández, P. J., and Lopez-Nicolas, A. (2001). How wide is the gap? an investigation of gender wage differences using quantile regression. *Empirical economics*, 26:149–167.

Granger, C. and Sin, C. (1997). Estimating and forecasting quantiles with asymmetric least squares. Technical report, Working Paper, University of California, San Diego.

Gu, Y. and Zou, H. (2016). High-dimensional generalizations of asymmetric least squares regression and their applications.

Hagemann, A. (2013). Robust spectral analysis.

Jiang, C., Jiang, M., Xu, Q., and Huang, X. (2017). Expectile regression neural network model with applications. *Neurocomputing*, 247:73–86.

Jiang, Y., Lin, F., and Zhou, Y. (2021). The kth power expectile regression. *Annals of the Institute of Statistical Mathematics*, 73:83–113.

Jones, M. C. (1994). Expectiles and m-quantiles are quantiles. *Statistics & Probability Letters*, 20(2):149–153.

Kley, T. (2016). Quantile-based spectral analysis in an object-oriented framework and a reference implementation in r: The quantspec package. *Journal of Statistical Software, Articles*, 70(3):1–27.

Knight, K. (1998). Limiting distributions for l 1 regression estimators under general conditions. *Annals of statistics*, pages 755–770.

Koenker, R. (2005). Quantile regression. *Cambridge University Press.*

Koenker, R. (2017). Quantile regression: 40 years on. *Annual review of economics*, 9:155–176.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Kouretas, G. P., Zarangas, L., et al. (2005). Conditional autoregressive value at risk by regression quantiles: Estimating market risk for major stock markets. Technical report.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

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

Li, T.-H. (2012a). Detection and estimation of hidden periodicity in asymmetric noise by using quantile periodogram. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),*vol 2, pp. 3969-3972.

Li, T.-H. (2012b). Quantile periodograms. *Journal of the American Statistical Association*, 107(498):765–776.

Li, T.-H. (2014). Quantile periodogram and time-dependent variance. *Journal of Time Series Analysis*, 35(4):322–340.

Li, T.-H. (2023). Quantile-frequency analysis and deep learning for signal classification. *Journal of Nondestructive Evaluation*, 42(2):40.

Maadooliat, M., Sun, Y., and Chen, T. (2018). Nonparametric collective spectral density estimation with an application to clustering the brain signals. *To appear in Statistics in Medicine.*

Machado, J. A. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465.

Martínez-Murcia, F. J., Ortiz, A., Morales-Ortega, R., López, P., Luque, J. L., Castillo-Barnes, D., Segovia, F., Illan, I. A., Ortega, J., Ramirez, J., et al. (2019). Periodogram connectivity of eeg signals for the detection of dyslexia. In *Understanding the Brain Function and Emotions: 8th International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2019, Almería, Spain, June 3–7, 2019, Proceedings, Part I 8*, pages 350–359. Springer.

Meziani, A., Medkour, T., and Djouani, K. (2020). Penalised quantile periodogram for spectral estimation. *Journal of Statistical Planning and Inference*, 207:86–98.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847.

Peligrad, M. and Utev, S. (1997). Central limit theorem for linear processes. *The Annals of Probability*, 25(1):443–456.

Polat, K. and Güneş, S. (2007). Classification of epileptiform eeg using a hybrid system based on decision tree classifier and fast fourier transform. *Applied Mathematics and Computation*, 187(2):1017–1026.

Schnabel, S. K. and PHC, E. (2009). Non-crossing smooth expectile curves. In *Proceedings of the 24th International Workshop on Statistical Modelling. Ithaca, NY, USA*, pages 330–36.

Sharif, A., Bhattacharya, M., Afshan, S., and Shahbaz, M. (2021). Disaggregated renewable energy sources in mitigating co2 emissions: new evidence from the usa using quantile regressions. *Environmental Science and Pollution Research*, 28(41):57582–57601.

Shumway, R. H. and Stoffer, D. S. (2016). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.

Waltrup, L. S., Sobotka, F., Kneib, T., and Kauermann, G. (2015). Expectile and quantile regression—david and goliath? *Statistical Modelling*, 15(5):433–456.

Xu, Q., Chen, L., Jiang, C., and Yu, K. (2020). Mixed data sampling expectile regression with applications to measuring financial risk. *Economic Modelling*, 91:469–486.

Xu, Q., Ding, X., Jiang, C., Yu, K., and Shi, L. (2021). An elastic-net penalized expectile regression with applications. *Journal of Applied Statistics*, 48(12):2205–2230.

Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: a nonparametric approach. *Journal of nonparametric statistics*, 6(2-3):273–292.

8 Appendix

8.1 Lemmas and Proofs

Lemma 1 (Expectile Regression). Given a time series $\{y_t : t = 1, \ldots, n\}$ of length n with cumulative distribution function $F(\cdot)$ and finite second moments, and regressors $\{\mathbf{x}_{nt} : t = 1, \ldots, n\}$ with $\mathbf{x}_{nt} := [x_{nt}(1), \ldots, x_{nt}(p)]^T$ and $x_{nt}(1) := 1$, Let the following assumptions be satisfied:

A1. The probability distribution function $F(\cdot)$ is Lipschitz continuous, i.e., there exists a constant $K > 0$ such that $|F(y) - F(y')| \leq K|y - y'|$ for all $y, y' \in \mathbb{R}$.

A2. The regressor sequence $\{\mathbf{x}_{nt}\}$ is bounded (in the ℓ_2-norm).

A3. There exists a positive-definite matrix $\mathbf{D}_0$ such that

$$\mathbf{D}_n := n^{-1} \sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{nt}^T \geq \mathbf{D}_0 \tag{18}$$

for sufficiently large n.

A4. The process $\{y_t\}$ has the strong mixing property with mixing numbers a_τ ($\tau = 1, 2, \ldots$) satisfying $a_\tau \to 0$ as $\tau \to \infty$.

Then, as $n \to \infty$, the ER solution given by (3) satisfies

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha)) = \boldsymbol{\Lambda}_n^{-1}(\alpha)\boldsymbol{\zeta}_n(\alpha) + o_P(1), \tag{19}$$

where $\boldsymbol{\beta}_0(\alpha) := [\mu(\alpha), 0, \ldots, 0]^T \in \mathbb{R}^p$, $\boldsymbol{\zeta}_n(\alpha)$ is given by (7), and

$$\boldsymbol{\Lambda}_n(\alpha) := \eta^{-1}(\alpha)\mathbf{D}_n. \tag{20}$$

The expression (19) is a Bahadur-type representation for the ER solution.

Proof. The proof follows the general strategy which was used to establish a similar result for quantile regression (Knight, 1998; Koenker, 2005; Li, 2008, 2012b).

Let $U_t := y_t - \mu(\alpha) = y_t - \mathbf{x}_{nt}^T \boldsymbol{\beta}_0(\alpha)$ and $c_{nt}(\boldsymbol{\delta}) := n^{-1/2}\mathbf{x}_{nt}^T\boldsymbol{\delta}$ with $\boldsymbol{\delta} \in \mathbb{R}^p$, consider the random function

$$Z_n(\boldsymbol{\delta}) := \sum_{t=1}^{n}\{\rho_\alpha(U_t - c_{nt}(\boldsymbol{\delta})) - \rho_\alpha(U_t)\}.$$

Note that by reparameterizing $\boldsymbol{\delta} := \sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0(\alpha))$ as a function of $\boldsymbol{\beta} \in \mathbb{R}^p$, one can write

$$Z_n(\boldsymbol{\delta}) = \sum_{t=1}^{n}\{\rho_\alpha(y_t - \mathbf{x}_{nt}^T\boldsymbol{\beta}) - \rho_\alpha(U_t)\}.$$

Therefore, the minimizer of $Z_n(\boldsymbol{\delta})$ over $\boldsymbol{\delta} \in \mathbb{R}^p$ takes the form

$$\hat{\boldsymbol{\delta}}_n := \arg\min_{\boldsymbol{\delta}\in\mathbb{R}^p} Z_n(\boldsymbol{\delta}) = \sqrt{n}(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha)).$$

It suffices to show

$$\hat{\boldsymbol{\delta}}_n = \boldsymbol{\Lambda}_n^{-1}\boldsymbol{\zeta}_n + o_P(1). \tag{21}$$

Here we drop the argument α from $\boldsymbol{\Lambda}_n$ and $\boldsymbol{\zeta}_n$ for simplicity of notation, as α is a fixed number. In order to arrive at (21), we first would like to prove that

$$Z_n(\boldsymbol{\delta}) = \tilde{Z}_n(\boldsymbol{\delta}) + o_P(1), \tag{22}$$

for any fixed $\boldsymbol{\delta}$, where $\tilde{Z}_n(\boldsymbol{\delta}) := -\boldsymbol{\delta}^T\boldsymbol{\zeta}_n + (1/2)\boldsymbol{\delta}^T\boldsymbol{\Lambda}_n\boldsymbol{\delta}$. If this is true, then, due to the convexity of $Z_n(\boldsymbol{\delta})$ and $\tilde{Z}_n(\boldsymbol{\delta})$ as functions of $\boldsymbol{\delta}$, one can follow the argument of Koenker (2005); Li (2008, 2012b) to show that $\hat{\boldsymbol{\delta}}_n$ as the minimizer of $Z_n(\boldsymbol{\delta})$ is $o_P(1)$ away from the minimizer of $\tilde{Z}_n(\boldsymbol{\delta})$, which equals $\boldsymbol{\Lambda}_n^{-1}\boldsymbol{\zeta}_n$. This completes the proof of (21).

To prove (22), we observe that for any $u, c \in \mathbb{R}$, we can write

$$\rho_\alpha(u-c) - \rho_\alpha(u) = \begin{cases} -2\alpha uc + \alpha c^2 & \text{if } u \ge 0 \text{ and } u - c \ge 0, \\ (1-2\alpha)u^2 - 2(1-\alpha)uc + (1-\alpha)c^2 & \text{if } u \ge 0 \text{ and } u - c < 0, \\ -(1-2\alpha)u^2 - 2\alpha uc + \alpha c^2 & \text{if } u < 0 \text{ and } u - c \ge 0, \\ -2(1-\alpha)uc + (1-\alpha)c^2 & \text{if } u < 0 \text{ and } u - c < 0. \end{cases}$$

$$
= \begin{cases} -\dot{\rho}_\alpha(u)c + \alpha c^2 & \text{if } u \geq 0 \text{ and } u - c \geq 0, \\ -\dot{\rho}_\alpha(u)c + (1-2\alpha)(u^2 - 2uc) + (1-\alpha)c^2 & \text{if } u \geq 0 \text{ and } u - c < 0, \\ -\dot{\rho}_\alpha(u)c - (1-2\alpha)(u^2 - 2uc) + \alpha c^2 & \text{if } u < 0 \text{ and } u - c \geq 0, \\ -\dot{\rho}_\alpha(u)c + (1-\alpha)c^2 & \text{if } u < 0 \text{ and } u - c < 0. \end{cases}
$$

Substituting U_t and $c_{nt} := c_{nt}(\boldsymbol{\delta})$ for u and c yields

$$
Z_n(\boldsymbol{\delta}) = -\boldsymbol{\delta}^T \boldsymbol{\zeta}_n + \sum_{t=1}^n R_{nt}, \tag{23}
$$

where

$$
R_{nt} := \begin{cases} \alpha c_{nt}^2 & \text{if } U_t \geq 0 \text{ and } U_t - c_{nt} \geq 0, \\ (1-2\alpha)(U_t^2 - 2U_t c_{nt}) + (1-\alpha)c_{nt}^2 & \text{if } U_t \geq 0 \text{ and } U_t - c_{nt} < 0, \\ -(1-2\alpha)(U_t^2 - 2U_t c_{nt}) + \alpha c_{nt}^2 & \text{if } U_t < 0 \text{ and } U_t - c_{nt} \geq 0, \\ (1-\alpha)c_{nt}^2 & \text{if } U_t < 0 \text{ and } U_t - c_{nt} < 0. \end{cases} \tag{24}
$$

Here, we drop the argument $\boldsymbol{\delta}$ from R_{nt} and c_{nt} for simplicity of notation as $\boldsymbol{\delta}$ is fixed in the remainder of the proof. Given (23), it suffices to show that

$$
\sum_{t=1}^n R_{nt} = \frac{1}{2}\boldsymbol{\delta}^T \boldsymbol{\Lambda}_n \boldsymbol{\delta} + o_P(1). \tag{25}
$$

This expression can be established by proving that

$$
\mathrm{E}\left\{\sum_{t=1}^n R_{nt}\right\} = \frac{1}{2}\boldsymbol{\delta}^T \boldsymbol{\Lambda}_n \boldsymbol{\delta} + o(1), \tag{26}
$$

and

$$
\mathrm{Var}\left\{\sum_{t=1}^n R_{nt}\right\} = o(1). \tag{27}
$$

To prove (26), we observe that

$$
\begin{aligned} \mathrm{E}\{R_{nt}\} = {} & \alpha c_{nt}^2 \mathrm{E}\{I(U_t \geq \max(0, c_{nt}))\} \\ & + \mathrm{E}\left\{[(1-2\alpha)(U_t^2 - 2U_t c_{nt}) + (1-\alpha)c_{nt}^2] I(0 \leq U_t < c_{nt})\right\} \\ & + \mathrm{E}\left\{[-(1-2\alpha)(U_t^2 - 2U_t c_{nt}) + \alpha c_{nt}^2] I(c_{nt} \leq U_t < 0)\right\} \\ & + (1-\alpha)c_{nt}^2 \mathrm{E}\{I(U_t < \min(0, c_{nt}))\}. \end{aligned} \tag{28}
$$

A necessary condition for the second term in (28) to be nonzero is $c_{nt} \geq 0$. When this is the case, the absolute value of this term can be bounded above by

$$
\left\{|1-2\alpha|(c_{nt}^2 + 2c_{nt}^2) + (1-\alpha)c_{nt}^2\right\} \Pr\{0 \leq U_t < c_{nt}\}.
$$

Under assumption (A2), $c_{nt} = o(1)$. These, together with assumption (A1), imply that

$$\Pr\{0 \leq U_t < c_{nt}\} = F(\mu(\alpha) + c_{nt}) - F(\mu(\alpha)) = o(1).$$

Therefore, the second term in (28) can be written as $o(c_{nt}^2)$. This expression is also valid for the third term in (28) which can be nonzero when $c_{nt} < 0$. Similarly, when $c_{nt} \geq 0$, the first term in (28) can be written as

$$\begin{aligned}
\alpha c_{nt}^2 \Pr\{U_t \geq c_{nt}\} &= \alpha c_{nt}^2 \Pr\{U_t \geq 0\} - \alpha c_{nt}^2 \Pr\{0 \leq U_t < c_{nt}\} \\
&= \alpha c_{nt}^2 \Pr\{U_t \geq 0\} + o(c_{nt}^2) \\
&= \alpha c_{nt}^2 \{1 - F(\mu(\alpha))\} + o(c_{nt}^2).
\end{aligned}$$

And, when $c_{nt} < 0$, the first term in (28) becomes simply $\alpha c_{nt}^2 \Pr\{U_t \geq 0\} = \alpha c_{nt}^2 \{1 - F(\mu(\alpha))\}$. By a similar argument, we can write the fourth term in (28) as

$$(1-\alpha) c_{nt}^2 F(\mu(\alpha)) + o(c_{nt}^2)$$

when $c_{nt} < 0$ and as $(1-\alpha) c_{nt}^2 F(\mu(\alpha))$ when $c_{nt} \geq 0$. Under assumption (A2), $o(c_{nt}^2) = o(n^{-1})$. Combining these results yields

$$\mathrm{E}\{R_{nt}\} = \{\alpha(1 - F(\mu(\alpha))) + (1-\alpha) F(\mu(\alpha))\} c_{nt}^2 + o(n^{-1}).$$

Substituting $c_{nt} = n^{-1/2} \mathbf{x}_{nt}^T \boldsymbol{\delta}$ in this expression proves (26).

Furthermore, for any $0 < m < n$, we split the variance in (27) into two terms:

$$\operatorname{Var}\left\{\sum_{t=1}^{n} R_{nt}\right\} = \left(\sum_{(t,s)\in D_m} + \sum_{(t,s)\in D_m'}\right) \operatorname{Cov}(R_{nt}, R_{ns}), \tag{29}$$

where $D_m := \{(t,s) : |t-s| \leq m, 1 \leq t, s \leq n\}$ and $D_m' := \{(t,s) : |t-s| > m, 1 \leq t, s \leq n\}$. The expression (27) is obtained if both terms in (29) can be shown to take the form $o(1)$.

Consider the first term in (29). Under assumption (A2), for any fixed $\boldsymbol{\delta}$, there exists a constant $c_0 := c_0(\boldsymbol{\delta}) > 0$ such that $c_{nt}^2 \leq c_0 n^{-1}$ for all $t = 1, \ldots, n$. Therefore, in the first and fourth cases of (24), we have, respectively, $|R_{nt}| = \alpha c_{nt}^2 \leq \alpha c_0 n^{-1}$ and $|R_{nt}| = (1-\alpha) c_{nt}^2 \leq (1-\alpha) c_0 n^{-1}$. If the second case of (24) is true, we would have $c_{nt} > 0$ and $0 \leq U_t < c_{nt}$, which implies

$$\begin{aligned}
|R_{nt}| &\leq |1 - 2\alpha|(c_{nt}^2 + 2c_{nt}^2) + (1-\alpha) c_{nt}^2 \\
&= (3|1 - 2\alpha| + (1-\alpha)) c_{nt}^2 \\
&\leq (3|1 - 2\alpha| + (1-\alpha)) c_0 n^{-1}.
\end{aligned}$$

Similarly, if the third case of (24) is true, we would have $c_{nt} < 0$ and $|U_t| \leq |c_{nt}|$, which implies

$$\begin{aligned}|R_{nt}| &\leq |1-2\alpha|(c_{nt}^2 + 2c_{nt}^2) + \alpha c_{nt}^2 \\ &= (3|1-2\alpha| + \alpha)c_{nt}^2 \\ &\leq (3|1-2\alpha| + \alpha)c_0\, n^{-1}.\end{aligned}$$

Combining these results yields

$$|R_{nt}| \leq \kappa\, n^{-1} \tag{30}$$

for some constant $\kappa := \kappa(\boldsymbol{\delta}, \alpha) > 0$. Owing to (30), we obtain

$$\mathrm{Var}\{R_{nt}\} \leq \mathrm{E}\{R_{nt}^2\} \leq \kappa^2 n^{-2}.$$

This implies that

$$|\mathrm{Cov}(R_{nt}, R_{ns})| \leq \sqrt{\mathrm{Var}\{R_{nt}\}\mathrm{Var}\{R_{ns}\}} \leq \kappa^2 n^{-2}. \tag{31}$$

Observe that the number of elements in D_m is bounded above by $(2m+1)n$. Combining this result with (31) yields

$$\left|\sum_{(t,s)\in D_m} \mathrm{Cov}(R_{nt}, R_{ns})\right| \leq \sum_{(t,s)\in D_m} |\mathrm{Cov}(R_{nt}, R_{ns})| \leq (2m+1)\kappa^2 n^{-1}.$$

If m is chosen to satisfy $m \to \infty$ and $m/n \to 0$ as $n \to \infty$, then we can write

$$\sum_{(t,s)\in D_m} \mathrm{Cov}(R_{nt}, R_{ns}) = o(1). \tag{32}$$

This completes the evaluation for the first term in (29).

To evaluate the second term in (29), we observe that $\{U_t\}$ is a stationary process with the same strong mixing property as $\{y_t\}$ under assumption (A4). Because R_{nt} is a function of U_t and R_{ns} is a function of U_s, both bounded by κn^{-1} according to (30), citing the mixing inequality (Billingsley, 2013)(Chen et al., 2021a) yields

$$|\mathrm{Cov}(R_{nt}, R_{ns})| \leq 4\kappa^2 n^{-2} a_{|t-s|}.$$

Therefore, we have

$$
\begin{aligned}
\left| \sum_{(t,s)\in D'_m} \mathrm{Cov}(R_{nt}, R_{ns}) \right| &\leq \sum_{m<|\tau|<n, 1\leq s\leq n} |\mathrm{Cov}(R_{n,s+\tau}, R_{ns})| \\
&\leq \sum_{m<|\tau|<n} n \times 4\kappa^2 n^{-2} a_{|\tau|} \\
&\leq 8\kappa^2 n^{-1} \sum_{\tau=m+1}^{n} a_\tau \\
&= 8\kappa^2 \left\{ n^{-1} \sum_{\tau=1}^{n} a_\tau - (m/n) m^{-1} \sum_{\tau=1}^{m} a_\tau \right\}.
\end{aligned}
$$

Because $a_\tau \to 0$ as $\tau \to \infty$, it follows from the Stolz-Cesàro theorem that $n^{-1}\sum_{\tau=1}^{n} a_\tau \to 0$ as $n \to \infty$ and $m^{-1}\sum_{\tau=1}^{m} a_\tau \to 0$ as $m \to \infty$. This implies that

$$
\sum_{(t,s)\in D'_m} \mathrm{Cov}(R_{nt}, R_{ns}) = o(1). \tag{33}
$$

Collecting (32) and (33) proves (27). Then, **Lemma 1** is proved.

Next, we would like to establish the asymptotic normality of the ER solution. Toward that end, consider $\boldsymbol{\zeta}_n(\alpha)$ defined by (7). Because $\mu(\alpha)$ satisfies (2), it follows that

$$
\mathrm{E}\{\boldsymbol{\zeta}_n(\alpha)\} = n^{-1/2} \sum_{t=1}^{n} \mathrm{E}\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\}\, \mathbf{x}_{nt} = \mathbf{0}.
$$

Moreover, we have

$$
\mathbf{V}_n(\alpha) := \mathrm{Cov}\{\boldsymbol{\zeta}_n(\alpha)\} = n^{-1} \sum_{t=1}^{n}\sum_{s=1}^{n} \mathrm{Cov}\{\dot{\rho}_\alpha(y_t - \mu(\alpha)), \dot{\rho}_\alpha(y_s - \mu(\alpha))\}\, \mathbf{x}_{nt}\mathbf{x}_{ns}^T. \tag{34}
$$

The following assertion is a direct result of Lemma 1.

Lemma 2 (Expectile Regression). Let the following assumptions be satisfied.

A5. There exists a positive definite matrix $\mathbf{D}$ such that $\mathbf{D}_n \to \mathbf{D}$ as $n \to \infty$.

A6. There exists a positive definite matrix $\mathbf{V}(\alpha)$ such that $\mathbf{V}_n(\alpha) \to \mathbf{V}(\alpha)$ as $n \to \infty$.

A7. A central limit theorem is valid for $\{\boldsymbol{\zeta}_n(\alpha)\}$, i.e., $\boldsymbol{\zeta}_n(\alpha) \xrightarrow{D} N(\mathbf{0}, \mathbf{V}(\alpha))$ as $n \to \infty$.

Then, under the additional assumptions (A1)-(A4) in **Lemma 1**, we have

$$
\sqrt{n}(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}_0(\alpha)) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Lambda}^{-1}(\alpha)\mathbf{V}(\alpha)\boldsymbol{\Lambda}^{-1}(\alpha)).
$$

where

$$\mathbf{\Lambda}(\alpha) := \lim_{n\to\infty} \mathbf{\Lambda}_n(\alpha) = \eta^{-1}(\alpha)\,\mathbf{D}.$$

Proof. Under assumptions (A5)-(A7), $\mathbf{\Lambda}_n(\alpha)^{-1}\boldsymbol{\zeta}_n(\alpha) \xrightarrow{D} N(\mathbf{0}, \mathbf{\Lambda}^{-1}(\alpha)\mathbf{V}(\alpha)\mathbf{\Lambda}^{-1}(\alpha))$. The assertion follows from (19) under assumptions (A1)-(A4). Then **Lemma 2** is proved.

For fixed $q > 1$, define $\{\mathbf{x}_{jnt} : t = 1, \ldots, n \text{ and } j = 1, \ldots q\}$ with $\mathbf{x}_{jnt} := [x_{jnt}(1), \ldots, x_{jnt}(p)]^T$ and $x_{jnt}(1) := 1$. Then, we extent **Lemma 2** with q regressors.

Lemma 2* (Expectile regression with multiple regressors). Let $\mathbf{D}_{jn}$ and $\mathbf{\Lambda}_{jn}(\alpha)$ take the form (18) and (20), except that $\mathbf{x}_{nt}$ is replaced by $\mathbf{x}_{jnt}$, respectively (hence we rewrite assumptions (A2) and (A3) into (A2*) and (A3*), respectively). $\boldsymbol{\zeta}_{jn}(\alpha)$ is denoted by (13). Moreover, we have

$$\mathbf{V}_{jkn}(\alpha) := n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathrm{Cov}\{\dot{\rho}_\alpha(y_t-\mu(\alpha)), \dot{\rho}_\alpha(y_s-\mu(\alpha))\}\,\mathbf{x}_{jnt}\mathbf{x}_{kns}^T, \quad j,k = 1,\ldots,q., \tag{35}$$

Then, we rewrite assumptions (A5), (A6), and (A7) with:

A5*. There exists a positive definite matrix $\mathbf{D}_j$ for all j such that $\mathbf{D}_{jn} \to \mathbf{D}_j$ as $n \to \infty$.

A6*. There exists a positive definite matrix $\mathbf{V}_{jk}(\alpha)$ for all j,k such that $\mathbf{V}_{jkn}(\alpha) \to \mathbf{V}_{jk}(\alpha)$ as $n \to \infty$.

A7*. A central limit theorem is valid for $\mathrm{vec}[\boldsymbol{\zeta}_{jn}(\alpha)]_{j=1}^q$ for all j, i.e., $\mathrm{vec}[\boldsymbol{\zeta}_{jn}(\alpha)]_{j=1}^q \xrightarrow{D} N(\mathbf{0}, [\mathbf{V}_{jk}(\alpha)]_{j,k=1}^q)$ as $n \to \infty$.

Then, if (A1), (A4), (A2*), (A3*), (A5*), (A6*), and (A7*) are satisfied, we have:

$$\sqrt{n}\mathrm{vec}[\hat{\boldsymbol{\beta}}_{jn}(\alpha) - \boldsymbol{\beta}_0(\alpha)]_{j=1}^q \xrightarrow{D} N(\mathbf{0}, [\mathbf{\Lambda}_j^{-1}(\alpha)\mathbf{V}_{jk}(\alpha)\mathbf{\Lambda}_k^{-1}(\alpha)]_{j,k=1}^q), \tag{36}$$

where

$$\widehat{\boldsymbol{\beta}}_{jn}(\alpha) := \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{t=1}^{n} \rho_\alpha(y_t - \mathbf{x}_{jnt}^T\boldsymbol{\beta}_j).$$

Proof. Similar to the proof of **Lemma 2**, define

$$Z_{jn}(\boldsymbol{\delta}_j) = \sum_{t=1}^{n}\{\rho_\alpha(y_t - \mathbf{x}_{jnt}^T\boldsymbol{\beta}_j) - \rho_\alpha(U_t)\},$$

and

$$Z_n^*(\boldsymbol{\delta}^*) = \sum_{j=1}^{q} Z_{jn}(\boldsymbol{\delta}_j),$$

where $\boldsymbol{\delta}^* := \mathrm{vec}[\boldsymbol{\delta}_j]_{j=1}^q$, and $\boldsymbol{\delta}_j := \sqrt{n}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0(\alpha))$. By a similar argument, it can be shown that the minimizer of $Z_n^*(\boldsymbol{\delta}^*)$, which can be expressed as $\hat{\boldsymbol{\delta}}_n^* := \mathrm{vec}[\hat{\boldsymbol{\delta}}_{jn}]_{j=1}^q$, is $o_P(1)$ away from $\tilde{\boldsymbol{\delta}}_n^* := \mathrm{vec}[\tilde{\boldsymbol{\delta}}_{jn}]_{j=1}^q := \mathrm{vec}[\boldsymbol{\Lambda}_{jn}^{-1}\boldsymbol{\zeta}_{jn}]_{j=1}^q$, where $\hat{\boldsymbol{\delta}}_{jn}$ is the minimizer of $Z_{jn}(\boldsymbol{\delta}_j)$. Therefore, $\hat{\boldsymbol{\delta}}_n^*$ has the same asymptotic distribution as $\tilde{\boldsymbol{\delta}}_n^*$. Observe that $\mathrm{E}\{\boldsymbol{\zeta}_{jn}\} = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\zeta}_{jn}, \boldsymbol{\zeta}_{kn}) = \mathbf{V}_{jkn}$. Therefore $\hat{\boldsymbol{\delta}}_n^* \xrightarrow{D} N(\mathbf{0}, [\boldsymbol{\Lambda}_j^{-1}(\alpha)\mathbf{V}_{jk}(\alpha)\boldsymbol{\Lambda}_k^{-1}(\alpha)]_{j,k=1}^q)$, and **Lemma 2*** is proved.

8.2 Proof of Theorem 1

For fixed $\omega \in (0, \pi)$, let $\omega_\nu := 2\pi v/n$ be a Fourier frequency such that $\omega_\nu \to \omega$ as $n \to \infty$ (i.e., let $\nu \to \infty$ such that $\nu/n \to \omega/(2\pi)$). This implies $\omega_\nu \in (0, \pi)$ for sufficiently large n. Under this condition, the trigonometric regressor $\mathbf{x}_{nt} := \mathbf{x}_t(\omega_\nu) := [1, \cos(\omega_\nu t), \sin(\omega_\nu t)]^T$ is bounded, so (A2) is satisfied. Let $\hat{\boldsymbol{\beta}}_n(\omega_\nu, \alpha)$ denote the ER solution given by (3). In this case, it is easy to verify that when $n \to \infty$,

$$\mathbf{D}_n = n^{-1}\sum_{t=1}^{n} \mathbf{x}_t(\omega_\nu)\mathbf{x}_t^T(\omega_\nu) \to \mathbf{D} = \mathrm{diag}\{1, 1/2, 1/2\}, \tag{37}$$

so (A3) and (A5) are satisfied. It follows from (34) that

$$\mathbf{V}_n(\alpha) = \sum_{|\tau|<n} \gamma(\tau, \alpha) \left\{ n^{-1} \sum_{t=\max(1,1+\tau)}^{\min(n,n+\tau)} \mathbf{x}_t(\omega_\nu)\mathbf{x}_{t-\tau}^T(\omega_\nu) \right\}.$$

For fixed τ, we have

$$n^{-1} \sum_{t=\max(1,1+\tau)}^{\min(n,n+\tau)} \mathbf{x}_t(\omega_\nu)\mathbf{x}_{t-\tau}^T(\omega_\nu) \to \mathrm{diag}\{1, (1/2)\mathbf{S}(\omega)\},$$

where

$$\mathbf{S}(\omega) := \begin{bmatrix} \cos(\omega\tau) & -\sin(\omega\tau) \\ \sin(\omega\tau) & \cos(\omega\tau) \end{bmatrix}.$$

Then,

$$\begin{aligned} \mathbf{V}(\alpha) &:= \lim_{n\to\infty} \mathbf{V}_n(\alpha) \\ &= \sum_{\tau=-\infty}^{\infty} \gamma(\tau, \alpha)\, \mathrm{diag}\{1, (1/2)\mathbf{S}(\omega)\} \\ &= \mathrm{diag}\{h(0, \alpha), (1/2)h(\omega, \alpha), (1/2)h(\omega, \alpha)\}, \end{aligned} \tag{38}$$

so (A6) is satisfied, and

$$\boldsymbol{\Lambda}^{-1}(\alpha)\mathbf{V}(\alpha)\boldsymbol{\Lambda}^{-1}(\alpha) = \mathrm{diag}\{g(0, \alpha), 2g(\omega, \alpha), 2g(\omega, \alpha)\}. \tag{39}$$

(C1), (C2) and (C3) repeat (A1), (A4) and (A7), respectively. Finally, all assumptions (A1)-(A7) are satisfied. Therefore, by **Lemma 1** and **Lemma 2**, (11) is proved, and the asymptotic variance satisfies (39).

Further, because $\hat{\boldsymbol{\beta}}_n(\omega_\nu, \alpha) = [\hat{\beta}_1(\omega_\nu, \alpha), \hat{\beta}_2(\omega_\nu, \alpha), \hat{\beta}_3(\omega_\nu, \alpha)]^T$, we have

$$\sqrt{n}[\hat{\beta}_2(\omega_\nu, \alpha), \hat{\beta}_3(\omega_\nu, \alpha)]^T / \sqrt{2g(\omega, \alpha)} \xrightarrow{D} N(\mathbf{0}, \mathbf{I}),$$

and hence

$$EP_n(\omega_\nu, \alpha)/g(\omega, \alpha) = \frac{n}{4}\{\hat{\beta}_2^2(\omega_\nu, \alpha) + \hat{\beta}_3^2(\omega_\nu, \alpha)\}/g(\omega, \alpha) \sim \frac{1}{2}(\xi_1^2 + \xi_2^2),$$

where ξ_1 and ξ_2 are i.i.d. $N(0,1)$ random variables. This leads to (12) with $\chi_2^2 := \xi_1^2 + \xi_2^2$.

Remark 2. Let $\hat{\beta}_1(0, \alpha)$ be the ER solution with $\mathbf{x}_{nt} := 1$. Then, under the assumption that the ACF $\gamma(\tau, \alpha)$ of $\{\dot{\rho}_\alpha(y_t - \mu(\alpha))\}$ is absolutely summable with $h(0, \alpha) > 0$, it follows from **Lemma 2** that

$$\sqrt{n}\{\hat{\beta}_1(0, \alpha) - \mu(\alpha)\} \xrightarrow{D} N(0, g(0, \alpha)).$$

Moreover, let $\hat{\boldsymbol{\beta}}_n(\pi, \alpha) := [\hat{\beta}_1(\pi, \alpha), \hat{\beta}_2(\pi, \alpha)]^T$ be the ER solution with $\mathbf{x}_{nt} := [1, \cos(\pi t)]^T = [1, (-1)^t]^T$. Then **Lemma 2** implies that

$$\sqrt{n}\{\hat{\boldsymbol{\beta}}_n(\pi, \alpha) - \boldsymbol{\beta}_0(\alpha)\} \xrightarrow{D} N(\mathbf{0}, \operatorname{diag}\{g(0, \alpha), g(\pi, \alpha)\}),$$

where $\boldsymbol{\beta}_0(\alpha) := [\mu(\alpha), 0]^T$. This means

$$EP_n(\pi, \alpha) = n\hat{\beta}_2^2(\pi, \alpha) \xrightarrow{D} g(\pi, \alpha)\chi_1^2,$$

where χ_1^2 denotes a random variable having the χ^2 distribution with 1 degree of freedom.

8.3 Proof of Theorem 2

This section is directly extended by the prove of **Theorem 1**. Under the conditions of **Theorem 2** with $\mathbf{x}_{jnt} := \mathbf{x}_{jt}(\omega_{\nu_j})$, all assumption of **Lemma 2*** are satisfied. Then, by (35),

$$\begin{aligned}\mathbf{V}_{jk}(\alpha) &:= \lim_{n\to\infty} \mathbf{V}_{jkn}(\alpha) \\ &= \delta_{j-k} \operatorname{diag}\{h(0, \alpha), (1/2)h(\lambda_j, \alpha), (1/2)h(\lambda_j, \alpha)\},\end{aligned}$$

where δ_s denotes the Kronecker delta function, and

$$[\boldsymbol{\Lambda}_j^{-1}(\alpha)\mathbf{V}_{jk}(\alpha)\boldsymbol{\Lambda}_k^{-1}(\alpha)]_{j,k=1}^q = \mathrm{diag}\{g(0,\alpha), 2g(\lambda_1,\alpha), 2g(\lambda_1,\alpha), \ldots, g(0,\alpha), 2g(\lambda_q,\alpha), 2g(\lambda_q,\alpha)\}.$$

By (36), we have

$$\sqrt{n}\mathrm{vec}[\hat{\boldsymbol{\beta}}_{jn}(\alpha)-\boldsymbol{\beta}_0(\alpha)]_{j=1}^q \xrightarrow{D} N(\mathbf{0}, \mathrm{diag}\{g(0,\alpha), 2g(\lambda_1,\alpha), 2g(\lambda_1,\alpha), \ldots, g(0,\alpha), 2g(\lambda_q,\alpha), 2g(\lambda_q,\alpha)\})$$

This implies that $EP_n(\omega_{\nu_j},\alpha)$ are asymptotically independent with asymptotic distribution $g(\lambda_j,\alpha)(1/2)\chi^2_{2,j}$ for each j, and we obtain (14).

8.4 Substitution of Assumption (A7)

Let us take a closer look at assumption (A7) (as well as (C3)). Using the Wald device, one can establish the central limit theorem in assumption (A7) by showing that

$$S_n := \boldsymbol{\lambda}^T\boldsymbol{\zeta}_n(\alpha)/\sqrt{\boldsymbol{\lambda}^T\mathbf{V}_n(\alpha)\boldsymbol{\lambda}} \xrightarrow{D} N(0,1).$$

for any fixed $\boldsymbol{\lambda} \neq \mathbf{0}$. Note that S_n can be written as the sum of a random sequence $\{w_{nt}\xi_t\}$, i.e.,

$$S_n = \sum_{t=1}^n w_{nt}\xi_t,$$

where $\xi_t := \dot{\rho}_\alpha(y_t - \mu(\alpha))$ and $w_{nt} := n^{-1/2}\boldsymbol{\lambda}^T\mathbf{x}_{nt}/\sqrt{\boldsymbol{\lambda}^T\mathbf{V}_n(\alpha)\boldsymbol{\lambda}}$. Note that $\mathrm{E}\{S_n\} = 0$ and $\mathrm{Var}\{S_n\} = 1$. A central theorem for this type of random variables is given by Peligrad and Utev (1997), where $\{\xi_t\}$ is assumed to be a possibly nonstationary random sequence with a strong mixing property. In our case, the sequence $\{\xi_t\}$ is stationary under assumption (A4). Therefore, Theorem 2.2(c) of Peligrad and Utev (1997) can be simplified as follows.

Proposition 1. (Peligrad and Utev, 1997). Let $\{\xi_t : t = 1, \ldots, n\}$ be a zero-mean nonzero-variance stationary sequence having the strong mixing property with mixing numbers a_τ ($\tau = 1, 2, \ldots$). Let $\{w_{nt} : t = 1, \ldots, n\}$ be a triangular array of real numbers. Assume that the following conditions are satisfied.

P1. For some $\delta > 0$, $\mathrm{E}\{|\xi_1|^{2+\delta}\} < \infty$ and $\sum_{\tau=1}^n \tau^{2/\delta}a_\tau < \infty$.

P2. $\sup_n \sum_{t=1}^n w_{nt}^2 < \infty$ and $\max_{1\le t\le n}|w_{nt}| \to 0$ as $n \to \infty$.

Under these conditions, $\sum_{t=1}^{n} w_{nt}\xi_t \xrightarrow{D} N(0,1)$ as $n \to \infty$.

Proof. When $\{\xi_t\}$ is stationary, we have $\inf_t \mathrm{Var}\{\xi_t\} = \mathrm{Var}\{\xi_1\} > 0$. If $\mathrm{E}\{|\xi_1|^{2+\delta}\} < \infty$, the stationarity also implies that for any $\epsilon > 0$ there exists $c > 0$ such that

$$\sup_t \mathrm{E}\{|\xi_t|^{2+\delta} I(|\xi_t| \geq c)\} = \mathrm{E}\{|\xi_1|^{2+\delta} I(|\xi_1| \geq c)\} < \epsilon.$$

This means that the sequence $\{|\xi_t|^{2+\delta}\}$ is uniformly integrable. Therefore, all conditions in Theorem 2.2(c) of Peligrad and Utev (1997) on $\{\xi_t\}$ are fulfilled under (P1) when $\{\xi_t\}$ is stationary. Condition (P2) repeats the assumption (2.1) in Peligrad and Utev (1997).

For the special case with $w_{nt} = n^{-1/2}\boldsymbol{\lambda}^T\mathbf{x}_{nt}/\sqrt{\boldsymbol{\lambda}^T\mathbf{V}_n(\alpha)\boldsymbol{\lambda}}$, condition (P2) is satisfied under the boundedness assumption (A2). Therefore, the following result can be obtained according to **Proposition 1**.

Lemma 3 (Expectile Regression). In addition to (A4), let $\{y_t\}$ also satisfy the following condition:

A8. For some $\delta > 0$, $\int |y|^{2+\delta}\, dF(y) < \infty$ and $\sum_{\tau=1}^{n} \tau^{2/\delta} a_\tau < \infty$.

Then, assumption (A7) holds if (A2) and (A6) are true.

Proof. Under (A4), $\xi_t := \dot{\rho}_\alpha(y_t - \mu(\alpha))$ is a zero-mean nonzero-variance strong mixing sequence with mixing numbers a_τ ($\tau = 1, 2, \ldots$). Under (A8),

$$\mathrm{E}\{|\xi_1|^{2+\delta}\} \leq \mathrm{E}\{|y_1|^{2+\delta}\} = \int |y|^{2+\delta}\, dF(y) < \infty.$$

Therefore, the assertion follows from the Wald device and **Proposition 1**. Then, one can substitute condition (C3) (also, A7) with (A8).