Evolutionary Spectra Based on the Multitaper Method with Application to Stationarity Test

Yu Xiang, *Member, IEEE,* Jie Ding, *Member, IEEE,* and Vahid Tarokh, *Fellow, IEEE*

Abstract—In this work, we propose a new inference procedure for understanding non-stationary processes, under the framework of evolutionary spectra developed by Priestley. Among various frameworks of modeling non-stationary processes, the distinguishing feature of the evolutionary spectra is its focus on the physical meaning of frequency. The classical estimate of the evolutionary spectral density is based on a double-window technique consisting of a short-time Fourier transform and a smoothing. However, smoothing is known to suffer from the so-called bias leakage problem. By incorporating Thomson's multitaper method that was originally designed for stationary processes, we propose an improved estimate of the evolutionary spectral density, and analyze its bias/variance/resolution tradeoff. As an application of the new estimate, we further propose a non-parametric rank-based stationarity test, and provide various experimental studies.

Index Terms—Non-stationary Processes, Evolutionary Spectra, Spectral Analysis, Multitaper Method, Stationarity Test.

I. Introduction

NOnstationary processes are common across a variety of areas and serve as a natural generalization of the classical wide-sense stationary processes. Because of their wide range of applications, they have been an active research area in many different areas including signal processing, statistics, neuroscience, and economics.

However, the intrinsic complexity of the non-stationarity precludes a unique way of modeling the non-stationary processes. Various frameworks have been developed over the past few decades: instantaneous power spectra [1], evolutionary spectra [2], Wigner-Ville spectral analysis [3], locally stationary processes [4], and local cosine basis [5] among others. In this work, we

adopt the *evolutionary spectra* framework developed by Priestley and his colleagues [2], [6]–[8], which is one of the first attempts to model non-stationary processes from the spectral point of view. The appealing aspect of this framework is its emphasis on the physical meaning of frequency, while generalizing the spectral representation of the stationary processes to that of the non-stationary processes [9].

Perhaps the most closely related framework is the *locally stationary processes* framework. Since Dahlhaus developed this framework in a series of papers [4], [10], [11], it has been extensively studied (see for example [12], [13] and references therein). We attempt to summarize the main differences between the two frameworks as follows. The evolutionary spectra framework is motivated by the physical interpretation of frequency, but does not guarantee the uniqueness of the spectral density. On the other hand, the locally stationary processes framework guarantees the uniqueness of the spectral density by providing an asymptotic analysis of the non-stationary processes. However, the rescaling technique, which is central to that framework, may sacrifice physical interpretations in some real applications. More comparisons between these two frameworks can also be found in [14] and detailed discussions on the other frameworks can be found in [9], [15] and the references therein.

The estimation procedure of the evolutionary spectra in [2] is based on the so-called double-window technique, consisting of a short-time Fourier transform and smoothing. However, the smoothing step is known to suffer from the so-called bias leakage problem [16]. To overcome this problem for stationary processes, various tapering methods have been developed and Thomson's *multitaper method* [17] is arguably the most widely used one. In this work, we apply the multitaper method to the estimation of evolutionary spectral density and analyze the bias/variance/resolution tradeoff of the estimate. We show that the non-stationarity calls for additional considerations of the tradeoff, which provides insights into window design, choice of frequency resolution and number of tapers. As an application of the estimate, we propose a non-parametric rank-based stationarity test

and compare it with the stationarity test investigated by Priestley and Subba Rao in [7]. Our test is more robust to the underlying distribution of the data, and it can serve as a complementary test to the existing stationarity tests from our numerical experiments.

The paper is organized as follows. In Section II, the evolutionary spectra framework is briefly reviewed. In Section III, the main results are summarized and the key steps of the proofs are presented. In Section IV, the estimate based on the short-time Fourier transform is evaluated through a new time-domain approach, which facilitates the analysis of the estimate in later sections and serves as a much simplified proof compared with the original one. In Section V, the estimate based on the multitaper method is analyze in the evolutionary spectra framework. In Section VI, a non-parametric stationarity test is proposed and various experimental studies are presented.

A. Notation

Let $\mathbb{Z}$ and $\mathbb{R}$ denote the set of integers and real numbers, respectively. For integers a and b such that $a < b$, let $[a : b] \triangleq \{a, a+1, \ldots, b\}$. Let $\{X(t)\} \triangleq \{X(t), t \in \mathbb{Z}\}$ denotes a sequence of random variables. $\{X(t)\}$ is called *wide-sense stationary* if its mean is a constant $\mathsf{E}[X(t)] = m_X$ and its auto-covariance depends only on the distance between the time indices $\mathrm{Cov}(X(t), X(s)) = c(t - s)$. Throughout this paper, stationary processes are referred to as wide-sense stationary processes. For two non-negative functions $f(x)$ and $g(x)$, we write $f(x) = O(g(x))$ if there exists some constant $0 < C < \infty$ such that $f(x) \leq Cg(x)$ for sufficiently large x. Let $\mathcal{N}(\mu, \sigma^2)$ denote a normal distribution with mean μ and variance σ^2. Let $\log(\cdot)$ denote the logarithm function with base 2. For $p \geq 1$, let $\|\cdot\|_p$ denote the l_p norm. We closely follow the notation in [18].

II. EVOLUTIONARY SPECTRA FRAMEWORK

A. Brief Review of the framework

In [2], the main focus is the continuous time setting, and the discrete time setting follows immediately. In this work, we will focus on the discrete time setting. In the following, we first briefly review the evolutionary spectra framework. Consider a class of non-stationary processes $\{X(t)\}$, with $\mathsf{E}[X(t)] = 0$ and $\mathsf{E}[X^2(t)] < \infty$ for $t \in \mathbb{Z}$, such that

$$X(t) = \int_{-\pi}^{\pi} \phi_t(w) dZ(w), t \in \mathbb{Z}, \qquad (1)$$

for some family $\mathcal{F}$ of functions $\{\phi_t(w)\}$ (defined on $[-\pi, \pi]$ indexed by t) and a measure $\mu(w)$, where $Z(w)$

is an orthogonal increment process with $\mathsf{E}|dZ(w)|^2 = d\mu(w)$. If there exists a family of functions $\{\phi_t(w) = e^{iwt}A_t(w)\}$ such that $\{X(t)\}$ can be represented as in (1) and any for fixed w, the Fourier transform $H_w(v)$ of $h_w(t) \triangleq A_t(w)$ (viewed as a function of t) has an absolute maximum at the origin, then $\{X(t)\}$ is called an *osciliatory process* with respect to *osciliatory functions* $\{e^{iwt}A_t(w)\}$, and the evolutionary spectrum at time t with respect to $\mathcal{F}$ is

$$dF_t(w) = |A_t(w)|^2 d\mu(w).$$

Remark 1: Note that $h_w(t) \equiv 1$ corresponds to the case when $\{X(t)\}$ is a stationary process, which leads to $H_w(v) \equiv \delta(v)$, where $\delta(\cdot)$ is the Dirac delta function.

Throughout this paper, we assume that $\mu(w)$ is absolutely continuous with respect to Lebesgue measure. Thus the *evolutionary spectral density at time t* is

$$f_t(w) = |A_t(w)|^2 \frac{d\mu(w)}{dw}.$$

Note that

$$H_w(v) = \sum_{t=-\infty}^{\infty} h_w(t) e^{-ivt}.$$

Without loss of generality, $A_t(w)$ can be normalized so that for all w,

$$A_0(w) = 1, \qquad (2)$$

which implies that $d\mu(w)$ represents the evolutionary spectrum at $t = 0$ and $|A_t(w)|^2$ represents the change relative to $t = 0$. Let

$$B_{\mathcal{F}}(w) = \int_{-\pi}^{\pi} |v||H_w(v)|dv,$$

and each family $\mathcal{F}$ of oscillatory functions is called *semi-stationary* if $B_{\mathcal{F}}(w)$ is bounded for all w. Then

$$B_{\mathcal{F}} = \left(\sup_w B_{\mathcal{F}}(w) \right)^{-1}$$

is call the *characteristic width of $\mathcal{F}$*. A *semi-stationary process* $\{X(t)\}$ is defined as the one that can be represented as (1) with respect to a semi-stationary family $\mathcal{F}$. Let $\mathcal{C}$ denote the class of semi-stationary families such that $\{X(t)\}$ can be represented as (1). Then

$$B_X = \sup_{\mathcal{F} \in \mathcal{C}} B_{\mathcal{F}} \qquad (3)$$

is called the *characteristic width of $\{X(t)\}$*. If there exists a family $\mathcal{F}^* \in \mathcal{C}$ with characteristic width equals to B_X, $\mathcal{F}^*$ is called the natural representation of $\{X(t)\}$. If there exists no family in $\mathcal{C}$ with characteristic width equals to B_X, let $\mathcal{F}^*$ denote any family with characteristic width arbitrarily close to B_X. From now on we will

only focus on $\mathcal{F}^*$ and the spectral representation with respect to this family. In particular, $A_t(w)$, $d\mu(w)$, and $dF_t(w)$ are all defined with respect to $\mathcal{F}^*$. In this work, we consider that $\{X(t)\}$ admits a natural representation $\mathcal{F}^*$, which implies that $B_{\mathcal{F}^*} = B_X$.

Remark 2: It is straightforward to see that semi-stationary processes includes stationary processes as special cases with $\mathcal{F}^* = \{e^{iwt}\}$ and $B_X = \infty$.

The characteristic width of $\{X(t)\}$, B_X, can be intuitively viewed as the maximal length of the interval over which the process may be treated as "approximately stationary" [2]. It plays an important role in this framework, however, it is hard to characterize (see Section V-B for a detailed discussion). Priestley [2] proposed a double-window technique to estimate the evolutionary spectral density. The first window is for the short-time Fourier transform and the second window is for smoothing. In this work, however, the second window will be replaced by the multitaper method as smoothing is known to suffer from the bias leakage problem (see Section V for details). The width of the first window $\{g(u), u \in \mathbb{R}\}$ is defined as

$$B_g \triangleq \sum_{u=-\infty}^{\infty} |u||g(u)|.$$

In this work, we focus on *time-limited* windows, i.e., there exists some $L > 0$ such that $g(u) = 0$ for $|u| > L$. Let $G(w)$ denote the Fourier transform of $g(u)$, i.e.,

$$G(w) = \sum_{u=-\infty}^{\infty} g(u)e^{-iuw}. \tag{4}$$

We assume that $g(u)$ is square integrable and without loss of generality it is normalized,

$$2\pi \sum_{u=-\infty}^{\infty} |g(u)|^2 = \int_{-\pi}^{\pi} |G(w)|^2 dw = 1. \tag{5}$$

B. Uniformly Modulated Processes

It is hard to characterize B_X exactly for semi-stationary processes [19]. However, there is one important class of processes whose characteristic widths can be bounded from below. This class, termed as the *uniformly modulated processes* [2], is of the following form:

$$X(t) = c(t)Y(t), \tag{6}$$

where $Y(t)$ is a stationary process with zero mean and spectral density $f_Y(w)$, and the Fourier transform of $c(t)$ has an absolute maximum at the origin. Thus it follows straightforwardly that

$$X(t) = \int_{-\pi}^{\pi} c(t)e^{iwt}dZ(w),$$

where $\mathsf{E}|dZ(w)|^2 = dF_Y(w)$. The process introduced in (6) is an oscillatory process since $\mathcal{F}_Y = \{c(t)e^{iwt}\}$ is a family of oscillatory functions. The evolutionary spectrum with respect to $\mathcal{F}$ is

$$f_t(w) = c^2(t)f_Y(w).$$

The name, uniformly modulated processes, follows from the fact that

$$\frac{f_{t_1}(w_1)}{f_{t_2}(w_1)} = \frac{f_{t_1}(w_2)}{f_{t_2}(w_2)}.$$

From the definition of B_X, we have $B_X \geq B_{\mathcal{F}_Y}$.

III. STATEMENTS OF THE MAIN RESULTS

For semi-stationary processes $\{X(t), 0 \leq t \leq T - 1\}$, it is natural to apply the *multitaper method* [17], which identifies K sequences of length N denoted by $\{g_k(u), 1 \leq k \leq K, 1 \leq u \leq N\}$ (assume N to be odd for simplicity of notation). Let $2\pi/N < W < \pi$ denote the frequency resolution of the multitaper method. The details are postponed to Section V. The estimate of the evolutionary spectral density $f_t(w)$ is given as below

$$\hat{f}_t^K(w) = \frac{1}{K} \sum_{k=0}^{K-1} |U^{(k)}(w)|^2, \tag{7}$$

where we have

$$U_t^{(k)}(w) = \sum_{u=0}^{T-1} g_k(u-t)X(u)e^{-iwu}$$

with $g_k(u)$ being a set of sequences (shifted so that they are centered around 0) each of length $N < T$ for $-(N-1)/2 \leq u \leq (N-1)/2$ and $1 \leq k \leq K$. Let

$$G_k(\lambda) = \sum_{u=-(N-1)/2}^{(N-1)/2} g_k(u)e^{-i\lambda u}.$$

The expectation of its evolutionary spectral density estimate using the multitaper method is given below.

Theorem 1:

$$\mathsf{E}[\hat{f}_t^K(w)]$$
$$= \int_{-\pi}^{\pi} \rho_K(w-\lambda)f_t(\lambda)d\lambda + O\big(B_g^{(K)}/B_X\big)$$
$$= \int_{-\pi}^{\pi} \rho_K(\lambda)f_t(w-\lambda)d\lambda + O\big(B_g^{(K)}/B_X\big),$$

where

$$\rho_K(\lambda) \triangleq \frac{1}{K} \sum_{k=0}^{K-1} |G_k(\lambda)|^2, \quad B_g^{(K)} \triangleq \max_k B_{g_k},$$

and $B_g^{(K)}$ is sufficiently smaller than B_X.

Assume that $\|f_t(w)\|_\infty$ is bounded for all t, the bias of the estimate can be bounded as follows.

Theorem 2: Assume that $\|f_t(w)\|_\infty < \infty$,

$$\left| \text{Bias}(\hat{f}_t^K(w)) \right| = \left| \hat{f}_t^K(w) - f_t^K(w) \right|$$
$$= O\left(\frac{\log N}{K} + W^2 + \frac{B_g^{(K)}}{B_X} \right).$$

When $\{X(t)\}$ is further assume to be a normal process, the variance of $\hat{f}_t^K(w)$ can be characterized as follows.

Theorem 3: Assume that $\|f_t(w)\|_\infty < \infty$ and $\{X(t)\}$ be a normal process,

$$\text{Var}(\hat{f}_t^K(w)) = O\left(\frac{1}{K} + \frac{B_g^{(K)}}{B_X} \right).$$

From Theorem 2 and 3, the mean squared error (MSE) of $\hat{f}_t^K(w)$ is given by the following.

Corollary 1: Given the same assumptions as in Theorem 3,

$$\text{MSE}(\hat{f}_t^K(w))$$
$$= O\left(\left(\frac{\log N}{K} \right)^2 + W^4 + \frac{1}{K} + \frac{B_g^{(K)}}{B_X} \right).$$

The proofs of the results are postponed to Section V and appendices and we briefly overview the main ingredients here. Firstly, we analyze $|U(w)|^2$ for general $\{g(u)\}$, which serves as a preliminary estimate (Proposition 1 and Proposition 2) before applying the multitaper method. We take a different approach than Priestley did in [2], in particular, we apply the *pseudo δ-function* argument (see Definition 1) directly in the time domain (Lemma 1) instead of in the frequency domain. The benefits of this new approach are twofolds, it makes the variance analysis for the multitaper method straightforward (Theorem 3) and provides a much simplified alternative proof of Proposition 2, which is a slightly different version of Theorem 8.1 in [2]. Secondly, by leveraging on a recent approximation result (Theorem 4) on multitaper method by Abreu and Romero [20], we analyze the bias/vairance/resolution tradeoffs in the evolutionary spectra framework.

IV. APPROXIMATELY UNBIASED ESTIMATE OF THE EVOLUTIONARY SPECTRA

In this section, we start with analyzing a preliminary estimate of the $f_t(w)$. Recall that $g(u)$ is assumed to be a time-limited function. First introduce $J_t(w)$ as follows, for fixed $t \in \mathbb{Z}$ and $w \in [-\pi, \pi]$,

$$J_t(w)$$

$$= \sum_{u=-\infty}^{\infty} g(u-t)X(u)e^{-iwu}$$

$$= \sum_{u=-\infty}^{\infty} g(u-t)\left(\int_{-\pi}^{\pi} A_u(\lambda)e^{i\lambda u}dZ(\lambda) \right)e^{-iwu}$$

$$\overset{(a)}{=} \int_{-\pi}^{\pi} \sum_{u=-\infty}^{\infty} g(u-t)A_u(\lambda)e^{-i(w-\lambda)u}dZ(\lambda)$$

$$= \int_{-\pi}^{\pi} e^{-i(w-\lambda)t} \sum_{v=-\infty}^{\infty} g(v)A_{v+t}(\lambda)e^{-i(w-\lambda)v}dZ(\lambda),$$

where the summation and integral can exchange in (a) is because $g(u)$ is a time-limited function.

Remark 3: Note that $J_t(w)$ is equivalent to its counterpart $Y_t(w)$ in [2] when $g(u)$ is a symmetric function, i.e., $g(u) = g(-u)$ for all u.

In the following, we introduce the pseudo δ-function argument but apply it to the time domain directly. The analysis of the spectra estimate of $f_t(w)$ in [2] depends on an approximation called *pseudo δ-function* and the discrete counterpart can be defined as below. The continuous version can be defined similarly and is used in [2].

Definition 1: Consider two functions $a(\cdot) : \mathbb{Z} \to \mathbb{R}$ and $b(\cdot) : \mathbb{Z} \to \mathbb{R}$. Then $a(u)$ is a pseudo δ-function of order ϵ with respect to $b(u)$ if, for any $t \in \mathbb{Z}$, there exists ϵ not depending on t such that

$$\left| \sum_{u=-\infty}^{\infty} a(u)b(u+t) - b(t) \sum_{u=-\infty}^{\infty} a(u) \right| < \epsilon.$$

Now we show the following.

Lemma 1: For family $\mathcal{F}^*$, $a(u) \triangleq g(u)e^{-iwu}$ is a pseudo δ-function of $b(u) \triangleq A_u(w)$ with order $O(B_g/B_X)$.

Proof: To clarify the role of u as the argument, we write $A_w(u) = A_u(w)$ in this proof. For any $t \in \mathbb{Z}$,

$$\sum_{u=-\infty}^{\infty} g(u)A_w(u+t)e^{-iwu}$$

$$= \sum_{u=-\infty}^{\infty} g(u)A_w(t)e^{-iwu} + R(t)$$

with

$$|R(t)|$$
$$\overset{(a)}{=} \frac{1}{2\pi}\left| \sum_{u=-\infty}^{\infty} g(u)e^{-iwu}\int_{-\pi}^{\pi} H_w(v)(e^{iv(t+u)} - e^{ivt})dv \right|$$

$$\leq \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} |g(u)|\int_{-\pi}^{\pi} |H_w(v)||e^{ivu} - 1|dv$$

$$\overset{(b)}{\leq} \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} |u||g(u)| \int_{-\pi}^{\pi} |v||H_w(v)|dv$$

$$\overset{(c)}{\leq} O(B_g/B_X),$$

where in (a), $A_w(u)$ is substituted by

$$A_w(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_w(v)e^{ivu}dv,$$

(b) follows since $|e^{ix} - 1| \leq |x|$, and (c) follows the definition of B_g and

$$\int_{-\pi}^{\pi} |v||H_w(v)|dv = B_{\mathcal{F}^*}(w) \leq 1/B_X.$$

∎

From Lemma 1, $J_t(w)$ can be further expressed in the following expression.

Proposition 1:

$$J_t(w)$$
$$= \int_{-\pi}^{\pi} e^{-i(w-\lambda)t}A_t(\lambda)\Big(\sum_{v=-\infty}^{\infty} g(v)e^{-i(w-\lambda)v}\Big)dZ(\lambda)$$
$$+ O(B_g/B_X) \int_{-\pi}^{\pi} e^{-i(w-\lambda)t}dZ(\lambda)$$
$$= \int_{-\pi}^{\pi} A_t(\lambda)G(w-\lambda)e^{-i(w-\lambda)t}dZ(\lambda)$$
$$+ O(B_g/B_X) \int_{-\pi}^{\pi} e^{-i(w-\lambda)t}dZ(\lambda), \quad (8)$$

where $G(w) = \sum_{v=-\infty}^{+\infty} g(v)e^{-iwv}$.

As one shall see, the relationship between the window choice $g(u)$ and the estimate $|J_t(w)|^2$ of $f_t(w)$ is revealed directly through this time domain approach. This leads to the following proposition.

Proposition 2:

$$\mathsf{E}[|J_t(w)|^2]$$
$$= \int_{-\pi}^{\pi} |G(w-\lambda)|^2 f_t(\lambda)d\lambda + O(B_g/B_X) \quad (9)$$
$$= \int_{-\pi}^{\pi} |G(\lambda)|^2 f_t(w-\lambda)d\lambda + O(B_g/B_X). \quad (10)$$

Proof: From (8) we have

$$\mathsf{E}[|J_t(w)|^2]$$
$$= \int_{-\pi}^{\pi} |G(w-\lambda)|^2 f_t(\lambda)d\lambda$$
$$+ O(B_g/B_X) \int_{-\pi}^{\pi} A_t(\lambda)G(w-\lambda)d\mu(\lambda) \quad (11)$$
$$+ O(B_g/B_X) \int_{-\pi}^{\pi} \overline{A_t(\lambda)G(w-\lambda)}d\mu(\lambda) \quad (12)$$

$$+ O\Big((B_g/B_X)^2\Big) \int_{-\pi}^{\pi} d\mu(\lambda). \quad (13)$$

Recall that

$$\int_{-\pi}^{\pi} |G(w)|^2 dw = 1.$$

Now (11) can be bounded as below.

$$\int_{-\pi}^{\pi} A_t(\lambda)G(w-\lambda)d\mu(\lambda)$$
$$\leq \int_{-\pi}^{\pi} |A_t(\lambda)||G(w-\lambda)|d\mu(\lambda)$$
$$\overset{(a)}{=} \int_{\Omega} |A_t(\lambda)||G(w-\lambda)|d\mu(\lambda)+$$
$$\int_{\Omega^c} |A_t(\lambda)||G(w-\lambda)|d\mu(\lambda)$$
$$\leq \int_{-\pi}^{\pi} d\mu(\lambda) + \int_{\Omega^c} |A_t(\lambda)|^2|G(w-\lambda)|^2 d\mu(\lambda)$$
$$\overset{(b)}{<} \infty,$$

where in (a) define $\Omega \triangleq \{w : |A_t(\lambda)||G(w-\lambda)| \leq 1\}$ and $\Omega^c \triangleq [-\pi, \pi]/\Omega$, (b) follows because of (5) and

$$\int_{\Omega^c} |A_t(\lambda)|^2|G(w-\lambda)|^2 d\mu(\lambda)$$
$$\leq \int_{-\pi}^{\pi} |G(w-\lambda)|^2 f_t(\lambda)d\lambda \leq \int_{-\pi}^{\pi} f_t(\lambda)d\lambda.$$

Since (12) and (13) can be bounded similarly, this ends the proof for (9). Observe that both $|G(w)|^2$ and $f_t(w)$ for fixed t are periodic functions with period 2π, thus (10) follows from (9). ∎

Given a sample record $\{X(0), X(1), ..., X(T-1)\}$ of length T, for $0 \leq t \leq T-1$, let

$$U_t(w) = \sum_{u=0}^{T-1} g(u-t)X(u)e^{-iwu}. \quad (14)$$

If we have that B_g is sufficiently smaller than B_X and B_X sufficiently smaller than T, then for t large enough, $U_t(w)$ becomes almost identical to $J_t(w)$ and the end effects are negligible. This holds since we are dealing with $g(u)$ that is time-limited, i.e., $g(u) = 0$ for $|u| > N$ for some N. Thus we have for $N/2 < t < T-N/2-1$,

$$\mathsf{E}[|U_t(w)|^2] = \int_{-\pi}^{\pi} |G(w-\lambda)|^2 f_t(\lambda)d\lambda + O(B_g/B_X).$$

To understand the impact of $|G(w)|^2$ on the bias of the estimate $|U_t(w)|^2$, consider the ideal case where $|G(w)|^2 = \delta(w)$. Then since $|G(w)|^2$ is normalized, we have

$$\mathsf{E}[|U_t(w)|^2] = f_t(w) + O(B_g/B_X).$$

Thus in this case $|U_t(w)|^2$ becomes an unbiased estimate of $f_t(w)$ up to $O(B_g/B_X)$. Intuitively speaking, the bias is controlled by the sidelobe of $|G(w)|^2$: less sidelobe would lead to a less biased estimate. To quantify the relationship, definitions similar to B_g and B_X in the frequency domain are needed. Let

$$\tilde{B}_g \triangleq \int_{-\pi}^{\pi} w|G(w)|^2 dw, \quad \tilde{B}_{\mathcal{F}}(t) \triangleq \int_{-\pi}^{\pi} |\lambda||L_t(\lambda)|d\lambda,$$

where $L_t(\lambda) \triangleq \int_{-\pi}^{\pi} f_t(v)e^{-i\lambda v} dv$. The minimum "width" of $f_t(w)$ for any t is thus $\tilde{B}_X \triangleq \sup_{\mathcal{F} \in \mathcal{C}} \tilde{B}_{\mathcal{F}}$ with $\tilde{B}_{\mathcal{F}} \triangleq \left(\sup_t \tilde{B}_{\mathcal{F}}(t)\right)^{-1}$.

Roughly speaking, $\tilde{B}_g$ and $\tilde{B}_X$ characterize the bandwidth of $g(u)$ and $f_t(w)$ in the frequency domain, respectively. Then in order to estimate $f_t(w)$, it has to be changing more slowly than $|G(w)|^2$ for each t. The following lemma is immediate and the proof can be found in Appendix A.

Lemma 2: $|G(w)|^2$ is a pseudo δ-function of $f_t(w)$ for each t with order $O(\tilde{B}_g/\tilde{B}_X)$.

Together with (5), Lemma 2 leads to

$$\mathsf{E}[|U_t(w)|^2] = f_t(w) + O(\tilde{B}_g/\tilde{B}_X) + O(B_g/B_X).$$

Therefore $|U_t(w)|^2$ is an unbiased estimate upto approximations in both time and frequency domain.

For stationary processes, it is well-known that simple periodogram type of estimate (as in (14) but without $\{g(u)\}$) produces Fejér kernel and is not a satisfactory estimate [18], [16]. In the asymptotic regime, either smoothing or tapering is needed to obtain a consistent estimate; while in the non-asymptotic regime, there is significant bias leakage because of the sidelobes of the Fejér kernel [16]. Different tapering techniques have been developed over the years and the *multitaper method* by Thomson [17] is the most widely used technique to reduce both the bias leakage and variance of the estimate. In the evolutionary spectra framework, however, additional constraint $O(B_g/B_X)$ is crucial to the performance of the estimate. Therefore the non-stationarity plays an important role in the bias/vairance/resolution tradeoff as we shall see in the next section.

V. ESTIMATE BASED ON THE MULTITAPER METHOD

Thomson's multitaper method [17] has been widely applied to various fields including wireless communincation [21], neuroscience [22], climate science [23]. In a recent paper by Abreu and Romero [20], the authors provide a rigorous proof of an important heuristic discovered by Thomson. They show that the averaged taper is close to an ideal band-pass filter over $[-W, W]$, i.e., $(1/2W)\mathbf{1}_{[-W,W]}$ in L_1 distance. First we briefly review Thomson's multitaper method [17] and the discrete prolate spheroidal sequences (DPSS) or Slepian sequences [24], [25], [16].

A. Thomson's Multitaper Method

Consider N sample records $\{X(0), \ldots, X(N-1)\}$ [1]. Assume that the sampling frequency is 1, then for a sequence of length N, the fundamental frequency is $2\pi/N$ and the Nyquist frequency is π. For $2\pi/N < W < \pi$, one wishes to find sequences with spectral densities concentrated over $[-W, W]$. We will refer to W as the resolution of the estimate. This problem was first investigated in a series of papers by Slepian, Laudau, Pollark [24]–[26]. The solution turns out to be a set of sequences $v_k(N, W; u), 0 \le u \le N-1, 0 \le k \le N-1$, which satisfy the following eigenvalue equation

$$\sum_{u'=0}^{N-1} \frac{\sin W(u - u')}{\sin \pi(u - u')} v_k(N, W; u')$$
$$= \lambda_k(N, W)v_k(N, W; u).$$

These N eigenvectors $v_k(N, W; \cdot)$ are called the discrete prolate spheroidal and they are ordered by their eigenvalues $1 > \lambda_0(N, W) > \lambda_1(N, W) > \cdots > \lambda_{N-1}(N, W) > 0$. It is well-known that the first $K = \lfloor 2NW/2\pi \rfloor = \lfloor NW/\pi \rfloor$ eigenvalues are close to 1.

Remark 4: The choice of using N as the length of the sample records instead of T is on purpose. T is the length of the whole sample records, while N will be used as the length of a time-limited function $g(u)$ discussed later in this section. If the process is indeed stationary, one would choose $T = N$.

The discrete prolate spheroidal wave functions are denoted by $V_k(N, W; \lambda)$ for $1 \le k \le K$, where

$$V_k(N, W; \lambda) = (-1)^k \epsilon_k \sum_{u=0}^{N-1} v_k(N, W; u)e^{-i\lambda(u-(N-1)/2)},$$

where $\epsilon_k = 1$ when k is even and $\epsilon_k = \sqrt{-1}$ when k is odd. For simplicity of notation, we suppress N and W and write $v_k(u) = v_k(N, W; u)$, $V_k(\lambda) = V_k(N, W; \lambda)$, and $\lambda_k = \lambda_k(N, K)$. These K functions satisfy two types of orthogonality over $[-W, W]$ and $[-\pi, \pi]$, respectively

$$\int_{-W}^{W} V_k(\lambda)\overline{V_l(\lambda)}d\lambda = \begin{cases} \lambda_k, & \text{for } k = l; \\ 0, & \text{otherwise.} \end{cases}$$

$$\int_{-\pi}^{\pi} V_k(\lambda)\overline{V_l(\lambda)}d\lambda = \begin{cases} 1, & \text{for } k = l; \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

[1] These N sample records are a consecutive subsequence from the $\{X(0), \ldots, X(T-1)\}$.

Observe that $|V_k(\lambda)|^2$ can be rewritten as follows,

$$|V_k(\lambda)|^2 = \left| \sum_{u=0}^{N-1} v_k(u) e^{-i\lambda u} \right|^2. \qquad (16)$$

Consider the average of the K tapered estimates,

$$\rho_K(\lambda) \triangleq \frac{1}{K} \sum_{k=0}^{K-1} |V_k(\lambda)|^2. \qquad (17)$$

It has been observed numerically that $\rho_K(\lambda)$ is close to $(1/2W)\mathbf{1}_{[-W,W]}(\cdot)$ by Thomson [17], which is justified recently by Abreu and Romero [20] as given below.

Theorem 4 ([20]): Let $N \geq 2$ denote the length of the sequence, $2\pi/N < W < \pi$ and set $K = \lfloor 2NW \rfloor$. Then

$$\left\| \rho_K(\cdot) - \frac{1}{2W} \mathbf{1}_{[-W,W]}(\cdot) \right\|_1 = O\left(\frac{\log N}{K} \right).$$

In the following section, we apply this result to analyze the performance of the multitaper method for semi-stationary processes.

B. Estimate of the Evolutionary Spectra based on the Multitaper Method

For stationary processes, the bias and variance of the multitaper spectral estimate (17) has been investigated [27], [28], [29]. In this section, we investigate its performance for semi-stationary processes. Let $g(u)$ be a time-limited function, i.e.,

$$|g(u)| = 0, \text{ for } |u| > (N-1)/2,$$

where N is assumed to be odd. Apply the multitaper method on $\{X(t), 0 \leq t \leq T-1\}$ with

$$g_k(u) \triangleq v_k\big(u + (N-1)/2\big) \text{ for } 0 \leq k \leq K-1,$$

then for $t > (N-1)/2$ we have

$$U_t^{(k)}(w) = \sum_{u=0}^{T-1} g_k(u-t) X(u) e^{-iwu}$$

$$= \sum_{u=t-(N-1)/2}^{t+(N-1)/2} g_k(u-t) X(u) e^{-iwu}.$$

From Proposition 2 and (16),

$$\mathsf{E}[|U_t^{(k)}(w)|^2]$$
$$= \int_{-\pi}^{\pi} |G_k(w-\lambda)|^2 f_t(\lambda) d\lambda + O\big(B_{g_k}/B_X\big),$$

where

$$G_k(\lambda) = \sum_{u=-\infty}^{\infty} g_k(u) e^{-i\lambda u} = \sum_{u=-(N-1)/2}^{(N-1)/2} g_k(u) e^{-i\lambda u}.$$

The estimate of $f_t(w)$ is the average of $|U_t^{(k)}(w)|^2$,

$$\hat{f}_t^K(w) = \frac{1}{K} \sum_{k=0}^{K-1} |U^{(k)}(w)|^2$$

and the mean of the estimate given in Theorem 1.

There is a bias/variance/resolution tradeoff for the estimate $\hat{f}_t^K(w)$. Assuming that $\|f_t(w)\|_\infty$ is bounded for all t, Theorem 2 can be proved by invoking Theorem 4 as given below.

Proof: First, the bias can be bounded,

$$|\text{Bias}(\hat{f}_t^K(w))|$$
$$= \left| \mathsf{E}[\hat{f}_t^K(w)] - f_t(w) \right|$$
$$\leq \left| \int_{-\pi}^{\pi} \rho_K(\lambda) f_t(w-\lambda) d\lambda - f_t(w) \right|$$
$$\qquad\qquad + O\big(B_g^{(K)}/B_X\big).$$

From Theorem 4,

$$\left| \int_{-\pi}^{\pi} \rho_K(\lambda) f_t(w-\lambda) d\lambda - f_t(w) \right|$$
$$\leq \left| \int_{-\pi}^{\pi} \big(\rho_K(\lambda) - \frac{1}{2W}\mathbf{1}_{[-W,W]}(\lambda)\big) f_t(w-\lambda) d\lambda \right|$$
$$\quad + \left| \int_{-\pi}^{\pi} \frac{1}{2W}\mathbf{1}_{[-W,W]}(\lambda) f_t(w-\lambda) d\lambda - f_t(w) \right|$$
$$\overset{(a)}{\leq} C(\log(N)/K + W^2),$$

where (a) follows from Theorem 4 and the assumption that $\|f_t(w)\|_\infty < \infty$ and C is some positive constant. $\blacksquare$

When $X(t)$ is a normal process, the variance of $\hat{f}_t^K(w)$ can be characterized as in Theorem 3 and the proof can be found in Appendix B. Now the MSE of $\hat{f}_t^K(w)$ can be bounded as shown in Corollary 1,

$$O\left(\left(\frac{\log N}{K}\right)^2 + W^4 + \frac{1}{K} + \frac{B_g^{(K)}}{B_X} \right),$$

where we have used the fact that the cross term $W^2(\log N/K)$ is dominated by either W^2 or $(\log N/K)$ depending on whether $W^2 \geq (\log N/K)$ or $W^2 \leq (\log N/K)$, and this argument applies to all the other cross terms as well.

For the uniformly modulated processes $X(t) = c(t)Y(t)$, where $Y(t)$ is a stationary process as defined in Section II-B, B_X can be lower bounded by $B_{\mathcal{F}_Y}$. Thus, based on Corollary 1, the MSE of the estimate in the case can be further bounded as

$$\left(\frac{\log N}{K}\right)^2 + W^4 + \frac{1}{K} + \frac{B_g^{(K)}}{B_{\mathcal{F}_Y}}.$$

Recall that $K = \lfloor 2NW \rfloor$ and W can be as small as $2\pi/N$. Thus for stationary processes, the MSE of the

spectral density estimate decreases as N grows. However, this is no longer the case for the semi-stationary processes as $B_g^{(K)}/B_X$ may become the dominant term for N large enough.

For non-stationary processes, it is a natural idea to approximate it by stationary processes locally. Heuristic methods, such as segmentation, have been developed [30] to deal with non-stationarity. In the evolutionary spectra framework, B_X defined in (3) can be roughly interpreted the longest "approximately stationary" segment [2] of the semi-stationary process. It is thus tempting to estimate B_X. However, it seems that B_X is more of a theoretical technique rather than providing fundamental meanings. Its definition is tailored to get the first order approximation of the estimate, which can be partially seen from Section IV. Furthermore, characterizing B_X is highly non-trivial as shown by Mélard in [19]. As a comparison, in the locally stationary processes framework [4], the authors characterized the optimal choice of N as N_{opt}, the length of the stationary segment [31] through minimizing the MSE of a local covariance estimate. While the characterization is interesting from the theoretical point of view, its application is limited due to its dependence on the true unknown parameters.

As a natural application of the evolutionary spectral density estimate, we propose a non-parametric stationarity test in the next Section.

VI. STATIONARITY TEST

The evolutionary spectral density estimate suggests a statistical test for the stationarity of a process, as first discussed in Priestley's paper [2] and later investigated by Priestley and Subba Rao (PSR test) in [7]. The original version of the PSR test uses the smoothing technique by introducing a second window, which suffers from bias leakage problems as discussed in Section IV. In a recent package developed by Constantine and Percival [32], smoothing is replaced by the multitaper method. This modified PSR test has been served as a baseline to when compared with other stationarity tests, e.g., in [33]. Based on the results from Section V-B, we attempt to provide some insights into the choice of the parameters in the test. Furthermore, a non-parametric version of the stationarity test is proposed, which is based on the Friedman test [34], [35] and is robust to the underlying distribution. It serves as a complementary test to the existing stationarity tests, in the sense that it is more conservative than PSR, see Section VI-C for details.

A. PSR Stationarity Test with the Multitaper method

Let $f_t(w)$ denote the evolutionary spectral density of a semi-stationary process $\{X(t) : 0 \leq t \leq T - 1\}$. Consider the estimate $\hat{f}_t^K(w)$ based on the multitaper method as in Section V and recall that

$$\hat{f}_t^K(w) = \frac{1}{K} \sum_{k=0}^{K-1} |U^{(k)}(w)|^2,$$

where

$$U_t^{(k)}(w) = \sum_{u=t-(N-1)/2}^{t+(N-1)/2} g_k(u-t)X(u)e^{-iwu}.$$

It is a common practice to take the logarithm of the estimate, which stabilizes its variance [36]. Let

$$Y_{ij} = \log \hat{f}_{t_i}^K(w_j),$$

Moreover, to apply the two-way analysis of variance (ANOVA) test [37], it has to be assumed that the distribution of $\log \hat{f}_t^K(w)$ is approximately normal [7]. More specifically, it can be shown that $W_{ij} = Y_{ij} - \psi(K) + \log(K)$ is approximately distributed according to the normal distribution with mean 0 and variance $\sigma^2 = \psi'(K)$ for $K \geq 5$, where $\psi(\cdot)$ and $\psi'(\cdot)$ denote the digamma function and the trigamma function, respectively (see [38] and [16] for details).

The approximate independence in time is obtained by choosing non-overlapping short windows of length N and the approximate independence in frequency is by choosing frequencies that are $2\pi(K+1)/(N+1)$ apart [2]. Now the problem reduces to a two-way ANOVA test for W_{ij} for $i \in [1 : I]$ and $j \in [1 : J]$, where $I = \lfloor T/N \rfloor$ and J is the number of frequencies chosen $2\pi(K+1)/(N+1)$ apart. Let

$$W_{..} = (1/IJ) \sum_{i=1}^{I} \sum_{j=1}^{J} W_{ij},$$

$$W_{i\cdot} = (1/J) \sum_{j=1}^{J} W_{ij},$$

$$W_{\cdot j} = (1/I) \sum_{i=1}^{I} W_{i\cdot}.$$

Between times variance with degrees of freedom $I - 1$ concerns how uniform are $\{W_{ij}\}$ over the time indices $1 \leq i \leq I$,

$$S_T = J \sum_{i=1}^{I} (W_{i\cdot} - W_{..})^2.$$

[2]Buffers are needed at the beginning and at the end when sample in frequency to overcome the edge effect. The size of the buffer could be chosen from $[B/2, B]$, where $B \triangleq 2\pi(K+1)/(N+1)$.

Similarly, between frequencies variance with degrees of freedom $J - 1$ is

$$S_F = I \sum_{j=1}^{J} (W_{\cdot j} - W_{\cdot\cdot})^2.$$

Interaction and residual variance with degrees of freedom $(I - 1)(J - 1)$ is

$$S_{I+R} = \sum_{i=1}^{I} \sum_{j=1}^{J} (W_{ij} - W_{i\cdot} - W_{\cdot j} + W_{\cdot\cdot})^2.$$

The null hypothesis is that the process is stationary and the alternative hypothesis is that the process is non-stationary. The test steps are described in the following.

1) First, test the interaction and residual sum of squares using $S_{I+R}/\sigma^2 = \chi^2_{(I-1)(J-1)}$.
2) If the interaction and residual is not significant, one conclude that is a uniformly modulated process. Then proceed to test $S_T/\sigma^2 = \chi^2_{(I-1)}$. If the between-times is not significant, conclude that the process is stationary. Otherwise, conclude that the process is non-stationary.
3) If the interaction and residual is significant, conclude that the process is non-stationary.

B. A Non-parametric Stationarity Test

There are two main assumptions of the two-way ANOVA test: (1) the samples are uncorrelated and (2) the residuals are normally distributed. There has been extensive research on the robustness of the assumptions for ANOVA test. It is known that the test statistics depend heavily on the first assumption and is less sensitive to the second assumption. The latter is shown empirically first in [39] and later in [40].

More specifically, the degree of violation of the normal distribution is usually characterized by the skewness β_1 and flatness β_2 of the distribution, where $\beta_1 = \mathsf{E}[(X - \mu)^3]/\sigma^3$ and $\beta_2 = \mathsf{E}[(X - \mu)^4]/\delta^4$, where μ and δ^2 denote the mean and variance of X, respectively. The test statistics are less sensitive to the skewness and flatness of X, essentially due to the central limit theorem as the the test statistics are based on summation of many terms. In the PSR test, the test results are more reliable when the degrees of freedom of time and frequency are large. On the other hand, nonparametric test, e.g., the rank-based Friedman test [34], [35], has an edge when the number of test samples is relatively small.

We now propose the non-parametric test, which will be referred to as *rank-based stationarity test* or *RS test* in short. Take $\{W_{ij}\}$ introduced in the previous section. In the time-frequency table filled by $\{W_{ij}\}$,

rank the elements in each column in an increasing order (i.e., 1 corresponds to the smallest element) to form a table of ranks: $\{R_{ij}\}$. Whenever there is a tie among k elements in the same column, assign the mean rank of the k elements. Similar to the two-way ANOVA test, let $R_{\cdot\cdot}$ denote the mean rank of all ranks, denote $R_{i\cdot}$ the mean rank of row i. The sum of square of ranks SS_R is $SS_R = J \sum_i (R_{i\cdot} - R_{\cdot\cdot})^2$. The test statistics $t_R = SS_R/\mathrm{const}$, where $\mathrm{const} = I(I + 1)/12$. It is known that t_R is (approximately) distributed according to χ^2_{I-1} [34], [35].

Remark 5: Conventionally, the rows are ranked and then the ranks in each column are summed up to form the test statistic. To be consistent with the two-way ANOVA test, the role of row and column are switched in this work.

C. Simulations

In this section, the performance of the proposed non-parametric stationarity test is evaluated and compared with the PSR test for a variety of synthetic data and real data. In our simulation, we use the multitaper function *pmtm* in MATLAB (R2016a) with default values as in [32]: number of tapers is 5, number of non-overlapping blocks is $\max\{2, \log(T)\}$, and buffer size $0.7B$ where $B = 2\pi(K + 1)/(N + 1)$.

1) Synthetic Data: The performance of a test is evaluated based on its empirical *size* and *power* values. Generate $M = 1000$ sample paths/realizations each with length $T = 512$ and let the nominal size of the test be 0.05. The null hypothesis H_0 is that the process is stationary and the alternative hypothesis H_1 is that the process is not stationarity.

For the size comparison, we generate sample paths from various stationary processes and count the number of rejections of the null hypothesis. Consider the following set of stationary autoregressive and moving-average (ARMA) models used in [41] The noise term $Z(t)$ is distributed according to $\mathcal{N}(0, 1)$.

(a) i.i.d. standard normal
(b) AR(1): $X(t) = 0.9X(t - 1) + Z(t)$.
(c) AR(1): $X(t) = -0.9X(t - 1) + Z(t)$.
(d) MA(1): $X(t) = Z(t) + 0.8Z(t - 1)$.
(e) MA(1): $X(t) = Z(t) - 0.8Z(t - 1)$.
(f) ARMA(1,2): $X(t) = -0.4X(t) + Z(t) - 0.8Z(t - 1)$.
(g) AR(2): $X(t) = \alpha_1 X(t-1) + \alpha_2 X(t-2) + Z(t)$ with $\alpha_1 = 1.385929$ and $\alpha_2 = -0.9604$ (from [42]).

The empirical sizes for PSR is smaller in [41] than that in Table I, but still at least twice as large as the empirical sizes of RS for all the models in Table I. The

Table I. Empirical size comparison (%)

models	PSR	RS
(a)	11.1	1.9
(b)	17.7	2.9
(c)	11.2	2.7
(d)	12.7	2.5
(e)	14.8	2.9
(f)	15.3	2.6
(g)	78.7	6.1

differences in PSR from [41] could be due to differences in parameters such as number of tapers, number of non-overlapping blocks, buffer size, etc.

For the power comparison, we generate sample paths from semi-stationary processes and count the number of acceptances of the null hypothesis. We focus on the uniformly modulated processes as in [2], [7]. The following model is from [7],

$$X(t) = e^{(t-T/2)^2/2a^2} Y(t), \qquad (18)$$

where $a = 200$ and $Y_t = 0.8Y_{t-1} - 0.4Y_{t-2} + Z_t$ with $Z_t \sim \mathcal{N}(0, 100^2)$. For all the models from Table I, generate uniformly modulated processes by multiplying each of them with $e^{(t-T/2)^2/2a^2}$. To make the numbering consistent with Table I, these models are also numbered from (a) to (g) and model (18) will be numbered as (h) in the table below.

Table II. Empirical power comparison (%)

models	PSR	RS
(a)	96.7	88.4
(b)	96.8	82.5
(c)	97.3	88.9
(d)	96.4	88.1
(e)	97.3	87.1
(f)	97.1	86.2
(g)	98.3	76.7
(h)	96.4	88.4

Since the empirical size of RS is smaller than that of PSR but the empirical power is also smaller, RS is a more conservative test compared with PSR.

2) Real Data: We consider a real data example called *ecgrr* used in [32] and comes from 'RR interval time series modeling: A challenge from PhysioNet and Computers in Cardiology 2002' site of PhysioNet. The data are the RR intervals (beat-to-beat intervals measured between successive peaks of the QRS complex) for patients in normal sinus rhythm (record 16265 of the

MIT-BIH database). The length of the sample $T = 512$ and the nominal size of the test be 0.05 as in the previous section. Recall that the test statistics for PSR are S_T/σ^2 and S_{I+R}/σ^2 and that for RS is t_R. Both PSR and RS suggest that the the process is non-stationary and their test statistics are summarized in the following tables.

Table III. Test result of PSR

	S_T/σ^2	S_{I+R}/σ^2
Chi-square quantile	16.919	72.1532
PSR	40.4927	67.7689

Table IV. Test result of RS

	t_R
Chi-square quantile	16.919
RS	22.8

3) Real Data: purchasing power parity: In economics, a common practice to test the purchasing power parity (PPP) hypothesis via testing the stationarity of real exchange rates (RER). Some earlier studies using unit root tests yielded results that were not favorable to PPP (see for example [43]–[45]). In this real data study, we test the stationarity of RER of four countries (Canada, China, Japan, UK) with respect to US over the period of January 1970 to December 2017. The monthly data of RER were calculated by $E \cdot P^*/P$, where E, P^* and P respectively denote the nominal exchange rates, the foreign price level (evaluated using consumer price index) and the domestic price level, using data sources from International Financial Statistics of the International Monetary Fund, Financial Statistics of the Federal Reserve Board, Haver Analytics, and the Pacific Exchange Rate Service. In the experiments, we take the widely used transform: the log first-order difference, i.e., $\log(X_t) - \log(X_{t-1})$ for $2 \le t \le N$ with $\{X_t : 1 \le N\}$ denote the RER.

Table V. Test results of PSR

	S_T/σ^2	S_{I+R}/σ^2
Chi-square quantile (5%)	15.5073	36.415
Chi-square quantile (1%)	20.0902	42.9798
PSR (Canada)	69.4236	42.6526
PSR (China)	637.9477	29.5066
PSR (Japan)	18.4479	13.6801
PSR (UK)	24.1467	32.9519

Based on RS test and S_T/σ^2 of the PSR test, one could order the stationarity of the four countries (in terms

Table VI. Tests result of RS

	t_R
Chi-square quantile (5%)	15.5073
Chi-square quantile (1%)	20.0902
RS (Canada)	22.0
RS (China)	30.8
RS (Japan)	16.3
RS (UK)	19.7

of RER) from the most stationary to least stationary as: Japan, UK, Canada, China. When choosing 1% as the nominal size of the test, RS test suggests that RER of both Japan and UK are stationary with Canada close to stationary; while PSR test suggests that RER of Japan is stationary and RER of UK is close to stationary. Both tests (under both 5% and 1%) suggest that RER of China is non-stationary.

APPENDIX A
PROOF OF LEMMA 1

Proof: The statement can be proved by observing

$$\int_{-\pi}^{\pi} |G(w)|^2 f_t(w+v)dw$$
$$= f_t(v) \int_{-\pi}^{\pi} |G(w)|^2 dw + R(v),$$

where

$$|R(v)| = |\int_{-\pi}^{\pi} w|G(w)|^2 f'_t(v+\eta(w)w)dw|$$
$$\leq \sup_{-\pi \leq w \leq \pi} |f'_t(w)| \int_{-\pi}^{\pi} |w||G(w)|^2 dw,$$

where $0 \leq \eta(w) \leq 1$ for all w. ∎

APPENDIX B
PROOF OF THEOREM 3

Proof: The variance of $\hat{f}_t^K(w)$ can be expressed as,

$$\text{Var}(\hat{f}_t^K(w))$$
$$= \text{Var}\left(\frac{1}{K} \sum_{k=0}^{K-1} |U_t^{(k)}(w)|^2\right)$$
$$= \frac{1}{K^2} \sum_{k,l=0}^{K-1} \text{Cov}\left(|U_t^{(k)}(w)|^2, |U_t^{(l)}(w)|^2\right).$$

First, $\text{Cov}\left(|U_t^{(k)}(w)|^2, |U_t^{(l)}(w)|^2\right)$ can be rewritten as (19), which is shown on the top of the next page. Similar to Proposition 2, we can further express S_1 and S_2

as (20) and (21), respectively. Now, since $\|f_t(w)\|_\infty < \infty$,

$$\left|\sum_{k,l=0}^{K-1} \text{Cov}\left(|U_t^{(k)}(w)|^2, |U_t^{(l)}(w)|^2\right)\right|$$
$$\leq \left|\sum_{k,l=0}^{K-1} S_1\right| + \left|\sum_{k,l=0}^{K-1} S_2\right|$$
$$\overset{(a)}{\leq} 2K + \sum_{k=0}^{K-1} O(B_{g_k}/B_X),$$

where (a) follows from (22) and a similar argument for S_2, as well as the orthogonality condition (15). This finishes the proof. ∎

REFERENCES

[1] C. H. Page, "Instantaneous power spectra," *Journal of Applied Physics*, vol. 23, no. 1, pp. 103–106, 1952.

[2] M. B. Priestley, "Evolutionary spectra and non-stationary processes," *Journal of the Royal Statistical Society. B*, vol. 28, no. 1, pp. 228–240, 1966.

[3] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 6, pp. 1461–1470, 1985.

[4] R. Dahlhaus, "On the Kullback-Leibler information divergence for locally stationary processes," *Stochastic Processes and their Applications*, vol. 62, pp. 139–168, 1996.

[5] S. Mallat, G. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *The Annals of Statistics*, vol. 26, no. 1, pp. 1–47, 1998.

[6] M. B. Priestley, "Design relations for non-stationary processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 228–240, 1966.

[7] M. B. Priestley and T. S. Rao, "A test for non-stationarity of time-series," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 140–149, 1969.

[8] M. B. Priestley and H. Tong, "On the analysis of bivariate non-stationary processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 153–166, 1973.

[9] M. B. Priestley, *Spectral analysis and time series*. Academic Press, 1981.

[10] R. Dahlhaus, "Fitting time series models to nonstationary processes," *The Annals of Statistics*, vol. 25, no. 1, pp. 1–37, 1997.

[11] ——, "A likelihood approximation for locally stationary processes," *The Annals of Statistics*, vol. 28, no. 6, pp. 1762–1794, 2000.

[12] G. P. Nason, R. Von Sachs, and G. Kroisandt, "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 2, pp. 271–292, 2000.

[13] H. C. Ombao, J. A. Raz, R. von Sachs, and B. A. Malow, "Automatic statistical analysis of bivariate nonstationary time series," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 543–560, 2001.

$$\mathrm{Cov}\bigg(|U_t^{(k)}(w)|^2, |U_t^{(l)}(w)|^2\bigg)$$

$$= \mathrm{Cov}\bigg(\sum_{u_1,u_2} g_k(u_1-t)g_k(u_2-t)X(u_1)X(u_2)e^{-iwu_1}e^{iwu_2}, \sum_{u_3,u_4} g_l(u_3-t)g_l(u_4-t)X(u_3)X(u_4)e^{-iwu_3}e^{iwu_4}\bigg)$$

$$\overset{(a)}{=} S_1 + S_2, \tag{19}$$

where (a) follows from the Isserlis' theorem [46] and $u_i \in [t-(N-1)/2 : t+(N-1)/2]$ for $i \in [1:4]$ with

$$S_1 = \sum_{u_1,u_2,u_3,u_4} g_k(u_1-t)g_k(u_2-t)g_l(u_3-t)g_l(u_4-t)\, \mathsf{E}[X(u_1)X(u_3)]\, \mathsf{E}[X(u_2)X(u_4)]e^{-iwu_1}e^{iwu_2}e^{-iwu_3}e^{iwu_4},$$

$$S_2 = \sum_{u_1,u_2,u_3,u_4} g_k(u_1-t)g_k(u_2-t)g_l(u_3-t)g_l(u_4-t)\, \mathsf{E}[X(u_1)X(u_4)]\, \mathsf{E}[X(u_2)X(u_3)]e^{-iwu_1}e^{iwu_2}e^{-iwu_3}e^{iwu_4}.$$

$$S_1 = \sum_{v_1,v_2,v_3,v_4} g_k(v_1)g_k(v_2)g_l(v_3)g_l(v_4)e^{-iw(v_1+t)}e^{iw(v_2+t)}e^{-iw(v_3+t)}e^{iw(v_4+t)}$$

$$\cdot \int_{-\pi}^{\pi} A_{v_1+t}(\lambda)A_{v_3+t}^*(\lambda)e^{i\lambda(v_1+t)}e^{-i\lambda(v_3+t)}d\mu(\lambda) \int_{-\pi}^{\pi} A_{v_2+t}(\xi)A_{v_4+t}^*(\xi)e^{i\xi(v_2+t)}e^{-i\xi(v_4+t)}d\mu(\xi)$$

$$= \int_{-\pi}^{\pi} \sum_{v_1,v_3} g_k(v_1)g_l(v_3)e^{-iw(v_1+t)}e^{-iw(v_3+t)}A_{v_1+t}(\lambda)A_{v_3+t}^*(\lambda)e^{i\lambda(v_1+t)}e^{-i\lambda(v_3+t)}d\mu(\lambda)$$

$$\int_{-\pi}^{\pi} \sum_{v_2,v_4} g_k(v_2)g_l(v_4)e^{iw(v_2+t)}e^{iw(v_4+t)}A_{v_2+t}(\xi)A_{v_4+t}^*(\xi)e^{i\xi(v_2+t)}e^{-i\xi(v_4+t)}d\mu(\xi)$$

$$\overset{(a)}{=} \int_{-\pi}^{\pi} |A_t(\lambda)|^2 \sum_{v_1,v_3} g_k(v_1)g_l(v_3)e^{-iw(v_1+t)}e^{-iw(v_3+t)}e^{i\lambda(v_1+t)}e^{-i\lambda(v_3+t)}d\mu(\lambda)$$

$$\int_{-\pi}^{\pi} |A_t(\xi)|^2 \sum_{v_2,v_4} g_k(v_2)g_l(v_4)e^{iw(v_2+t)}e^{iw(v_4+t)}e^{i\xi(v_2+t)}e^{-i\xi(v_4+t)}d\mu(\xi) + O(B_{g_k}/B_X) + O(B_{g_l}/B_X)$$

$$= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} G_k(w-\lambda)G_l(w+\lambda)G_k(-w-\xi)G_l(-w+\xi)f_t(\lambda)d\lambda f_t(\xi)d\xi + O(B_{g_k}/B_X) + O(B_{g_l}/B_X), \tag{20}$$

where (a) follows from Lemma 1. Similarly we have

$$S_2 = \sum_{v_1,v_2,v_3,v_4} g_k(v_1)g_k(v_2)g_l(v_3)g_l(v_4)e^{-iw(v_1+t)}e^{iw(v_2+t)}e^{-iw(v_3+t)}e^{iw(v_4+t)}$$

$$\cdot \int_{-\pi}^{\pi} A_{v_1+t}(\lambda)A_{v_4+t}^*(\lambda)e^{i\lambda(v_1+t)}e^{-i\lambda(v_4+t)}d\mu(\lambda) \int_{-\pi}^{\pi} A_{v_2+t}(\xi)A_{v_3+t}^*(\xi)e^{i\xi(v_2+t)}e^{-i\xi(v_3+t)}d\mu(\xi)$$

$$= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} G_k(w-\lambda)G_l(-w+\lambda)G_k(-w-\xi)G_l(w+\xi)f_t(\lambda)d\lambda f_t(\xi)d\xi + O(B_{g_k}/B_X) + O(B_{g_l}/B_X). \tag{21}$$

---

[14] R. Dahlhaus, "Asymptotic statistical inference for nonstationary processes with evolutionary spectra," *Athens Conference on Applied Probability and Time Series*, vol. II, pp. 145–159, 1996.

[15] P. Flandrin, *Time-frequency/time-scale analysis*. Academic press, 1998, vol. 10.

[16] D. B. Percival and A. T. Walden, *Spectral analysis for physical applications*. Cambridge University Press, 1993.

[17] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

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

[19] G. Mélard and A. H.-d. Schutter, "Contributions to evolutionary spectral theory," *Journal of Time Series Analysis*, vol. 10, no. 1, pp. 41–63, 1989.

[20] L. D. Abreu and J. L. Romero, "MSE estimates for multitaper spectral estimation and off-grid compressive sensing," *arXiv*

$$\sum_{k,l=0}^{K-1} S_1 = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{k=0}^{K-1} G_k(w-\lambda)G_k(-w-\xi) \sum_{l=0}^{K-1} G_l(w+\lambda)G_l(-w+\xi)f_t(\lambda)d\lambda f_t(\xi)d\xi + \sum_{k=0}^{K-1} O(B_{g_k}/B_X)$$

$$\leq \left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left| \sum_{k=0}^{K-1} G_k(w-\lambda)G_k(-w-\xi) \right|^2 f_t(\lambda)d\lambda f_t(\xi)d\xi \right.$$

$$\left. \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left| \sum_{l=0}^{K-1} G_l(w+\lambda)G_l(-w+\xi) \right|^2 f_t(\lambda)d\lambda f_t(\xi)d\xi \right)^{1/2} + \sum_{k=0}^{K-1} O(B_{g_k}/B_X). \tag{22}$$

preprint arXiv:1703.08190, 2017.

[21] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE journal on selected areas in communications*, vol. 23, no. 2, pp. 201–220, 2005.

[22] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[23] G. Bond, W. Showers, M. Cheseby, R. Lotti, P. Almasi, P. Priore, H. Cullen, I. Hajdas, G. Bonani *et al.*, "A pervasive millennial-scale cycle in North Atlantic holocene and glacial climates," *science*, vol. 278, no. 5341, pp. 1257–1266, 1997.

[24] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertaintyi," *Bell Labs Technical Journal*, vol. 40, no. 1, pp. 43–63, 1961.

[25] D. Slepian, "Prolate spheroidal wave functions, fourier analysis, and uncertaintyv: The discrete case," *Bell Labs Technical Journal*, vol. 57, no. 5, pp. 1371–1430, 1978.

[26] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertaintyii," *Bell Labs Technical Journal*, vol. 40, no. 1, pp. 65–84, 1961.

[27] K. Lii and M. Rosenblatt, "Prolate spheroidal spectral estimates," *Statistics & Probability Letters*, vol. 78, no. 11, pp. 1339–1348, 2008.

[28] A. T. Walden, E. McCoy, and D. B. Percival, "The variance of multitaper spectrum estimates for real Gaussian processes," *IEEE Transactions on Signal Processing*, vol. 42, no. 2, pp. 479–482, 1994.

[29] J. A. Hogan and J. D. Lakey, *Duration and bandwidth limiting: prolate functions, sampling, and applications*. Springer Science & Business Media, 2011.

[30] M. Ding, S. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment," *Biological Cybernetics*, vol. 83, no. 1, pp. 35–45, 2000.

[31] R. Dahlhaus and L. Giraitis, "On the optimal segment length for parameter estimates for locally stationary time series," *Journal of Time Series Analysis*, vol. 19, no. 6, pp. 145–159, 1998.

[32] W. Constantine and D. Percival, "Fractal: fractal time series modeling and analysis," *R package version*, 2011.

[33] G. Nason, "A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally

stationary time series," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 75, no. 5, pp. 879–904, 2013.

[34] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the american statistical association*, vol. 32, no. 200, pp. 675–701, 1937.

[35] ——, "A comparison of alternative tests of significance for the problem of m rankings," *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.

[36] U. Grenander and M. Rosenblatt, "Statistical analysis of stationary time series," 1957.

[37] F. Yates, "The analysis of multiple classifications with unequal numbers in the different classes," *Journal of the American Statistical Association*, vol. 29, no. 185, pp. 51–66, 1934.

[38] M. S. Bartlett and D. Kendall, "The statistical analysis of variance-heterogeneity and the logarithmic transformation," *Supplement to the Journal of the Royal Statistical Society*, vol. 8, no. 1, pp. 128–138, 1946.

[39] E. S. Pearson, "The analysis of variance in cases of non-normal variation," *Biometrika*, pp. 114–133, 1931.

[40] P. P. D. Glass, Gene V and J. R. Sanders, "Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance," *Review of educational research*, vol. 42, no. 3, pp. 237–288, 1972.

[41] G. Nason, "Simulation study comparing two tests of second-order stationarity and confidence intervals for localized autocovariance," *arXiv:1603.06415*, 2016.

[42] C. Chatfield, *The analysis of time series: an introduction*. CRC press, 2016.

[43] W. Enders, "Arima and cointegration tests of ppp under fixed and flexible exchange rate regimes," *The Review of Economics and Statistics*, pp. 504–508, 1988.

[44] M. P. Taylor, "An empirical examination of long-run purchasing power parity using cointegration techniques," *Applied economics*, vol. 20, no. 10, pp. 1369–1381, 1988.

[45] W. J. Crowder, "Testing stationarity of real exchange rates using johansen tests," *Department of Economics, University of Texas at Arlington*, 2001.

[46] L. Isserlis, "On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression," *Biometrika*, vol. 11, no. 3, pp. 185–190, 1916.