Journal of Econometrics 182 (2014) 211–225

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Spectral density and spectral distribution inference for long memory time series via fixed-b asymptotics

Tucker S. McElroy [a,1], Dimitris N. Politis [b,*]

[a] Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, USA

[b] Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, Mail Code 0112, La Jolla, CA 92093-0112, USA

ARTICLE INFO

Article history:
Available online 24 April 2014

JEL classification:
C1

Keywords:
Cyclical long memory
Kernel spectral estimator
Long range dependence
Spectral confidence bands

ABSTRACT

This paper studies taper-based estimates of the spectral density utilizing a fixed bandwidth ratio asymptotic framework, and makes several theoretical contributions: (i) we treat multiple frequencies jointly, (ii) we allow for long-range dependence or anti-persistence at differing frequencies, (iii) we allow for tapers that are only piecewise smooth or discontinuous, including flat-top and truncation tapers, (iv) we study higher-order accuracy through the limit distribution's Laplace Transform, (v) we develop a taper-based estimation theory for the spectral distribution, and show how confidence bands can be constructed. Simulation results produce quantiles and document the finite-sample size properties of the estimators, and a few empirical applications demonstrate the utility of the new methods.

Published by Elsevier B.V.

1. Introduction

Suppose that we have a sample $Y_1, Y_2, \ldots, Y_N$ from a weakly stationary time series $\{Y_t\}$, and consider a kernel-based estimator of the spectral density $f(\theta)$ defined via

$$\widehat{f}(\theta) = \sum_h \Lambda(h/M)\cos(\theta h)\widehat{\gamma}_h \tag{1}$$

for any fixed $\theta \in [-\pi, \pi]$. Here Λ is the kernel, or taper, and is a bounded even function of domain $[-1, 1]$. The sequence $\widehat{\gamma}_h$ consists of sample autocovariances, where the centering can be taken as either zero, the sample mean, or OLS estimates of a more complicated regression effect. The bandwidth M is taken to grow at the same rate as the sample size N, rather than the usual $o(N)$ growth rate, such that $M = bN$ for some $b \in (0, 1)$; we say that the bandwidth-ratio b is fixed, and use the terminology of fixed-b asymptotics. The following result is a consequence of Theorem 1 of Hashimzade and Vogelsang (2008) under assumptions consistent with a short memory time series:

$$\widehat{f}(\theta) \stackrel{\mathcal{L}}{\Longrightarrow} f(\theta) \cdot S_\theta(b)$$

as $N \to \infty$. The limiting random variable $S_\theta(b)$ is a quadratic functional of Brownian Motion that depends on the bandwidth proportion b, but not on the short memory autocorrelation function of the data process, and thus can be simulated without any knowledge of nuisance parameters. The limit also depends on the taper Λ, and the distribution depends on θ as well, since results differ depending on whether $\theta = 0$, $\theta = \pi$, or $\theta \in (0, \pi)$. Furthermore, the distribution at frequency $\theta = 0$ also depends on the type of centering used to define $\widehat{\gamma}_h$.

As noted in Hashimzade and Vogelsang (2008) – henceforth HV – the asymptotic coverage provided by the so-called large-bandwidth approach is superior when b is greater than zero, and also has the advantage of guaranteeing a positive random limit (when the taper Λ is positive definite). The potential application of a better inferential methodology for the spectral density function is quite large, as demonstrated by the ubiquity of spectral methods in the physical sciences as well as econometrics; see Grenander and Rosenblatt (1953), Parzen (1957), Blackman and Tukey (1959), Bohman (1960), and the discussion in Priestley (1981). Understanding the joint distribution of spectral estimates across multiple frequencies is useful for the identification of hidden periodicities in the time series. One application is the identification of residual seasonality in seasonally adjusted economic time series via examination of spectral estimates in the program X-12-ARIMA, as discussed in Findley et al. (1998). Literally millions of time series are seasonally adjusted each month by the program X-12-ARIMA at statistical agencies around the world – with vast ramifications for public policy – and spectral peak estimation and assessment is featured as a diagnostic tool in every application. Some other econometric applications of spectral estimation are discussed at the end of the paper.

* Corresponding author.
E-mail addresses: tucker.s.mcelroy@census.gov (T.S. McElroy), dpolitis@ucsd.edu (D.N. Politis).
[1] Tel.: +1 301 763-3227; fax: +1 301 763-8399.

http://dx.doi.org/10.1016/j.jeconom.2014.04.019
0304-4076/Published by Elsevier B.V.

The paper at hand seeks to make several extensions of the fundamental results of HV. First, we extend their basic results to a joint theorem over a finite collection of frequencies. This is important for assessing the uncertainty in taper-smoothed estimates of the spectral density, where we may be interested in 30–60 ordinates at a time. As our results below demonstrate, $S_{\theta_1}(b_1)$ is asymptotically independent of $S_{\theta_2}(b_2)$ for $\theta_1 \neq \theta_2$ and any $b_1, b_2 \in (0, 1]$. This technical result will allow us to construct simultaneous confidence intervals, allowing one to assess uncertainty in a nonparametric spectral analysis.

Second, we study cyclical long-range dependence, where each frequency of the spectral density may correspond to a long memory pole or a negative memory zero; see Boutahar (2008) for related asymptotic results for the case of a single frequency. Cyclical long memory is useful for capturing highly persistent seasonal or cyclical phenomena that evolve too rapidly to be considered nonstationary; see Holan and McElroy (2012) for examples and applications of the concept to the problem of seasonal adjustment. The presence of cyclical long memory implies that the rate of convergence of the spectral estimates depends on the corresponding memory parameter, and the limit distributions become quadratic functionals of Fractional Brownian Motion — this is an extension of the frequency zero results of McElroy and Politis (2012). The rate of growth of the spectral estimates is non-standard in this case, so that the resulting confidence intervals are much wider (for long memory) or shorter (for negative memory) than in the regular short memory scenario.

Third, we extend the limit theorems to piecewise smooth tapers, such as flat-top tapers (see Politis and Romano (1995) and Politis (2001)), and also to tapers with jump discontinuities, such as the truncation taper. With the exception of the Bartlett taper, HV and other fixed-b literature consider only smooth tapers (such as Parzen or Tukey–Hanning). For example, Phillips et al. (2006) derives asymptotic results for spectral estimates (handling multivariate time series) computed from smooth tapers, examining one frequency at a time. However, some popular tapers (such as Daniell and Quadratic-Spectral) have kinks (i.e., where a continuous function is not differentiable) at the boundary of their domain, which has an impact on the limit distribution—this was established in McElroy and Politis (2012) for the frequency zero case. Flat-top tapers have proven useful for variance estimation of short memory processes, so it seems important to develop the spectral theory for such tapers.

Fourth, we provide a discussion of higher-order accuracy of the limit theory arising from the fixed-bandwidth ratio methodology. In the recent literature on Heteroskedasticity–Autocorrelation Consistent (HAC) testing – see Kiefer et al. (2000) and Kiefer and Vogelsang (2002) – this has meant an expansion of the fixed-bandwidth ratio limit distributions as b tends to zero, such that the first term in the expansion is the conventional limit distribution of the vanishing-bandwidth ratio theory (i.e., in the HAC case a standard normal). We are not aware that a higher-order accuracy limit theory has been published for fixed bandwidth ratio spectral density estimates, though Velasco and Robinson (2001) study the vanishing bandwidth ratio case. Actually, the HAC literature shows that $S_0(b)$ tends to a point mass at unity as b tends to zero; correspondingly, the higher-order accuracy results in this paper demonstrate that the cumulative distribution function of $S_\theta(b)$ can likewise be expanded as $b \to 0$, with a leading term equal to an indicator function, followed by other expressions involving cumulants. To achieve this, we introduce a novel method of inverting the Laplace Transform of Gaussian quadratic forms.

It may be of some interest to provide a confidence band for the entire spectral density. This is not possible if long-range dependence is present, because each frequency would potentially be growing at different rates. Also, because the spectral density limit distributions across frequencies are independent in a fixed bandwidth ratio approach, the global behavior is better summarized through the spectral distribution function (Woodroofe and Van Ness (1967) consider the spectral density bands under a vanishing bandwidth fraction asymptotic approach). Although previous literature explores the estimation of the spectral distribution function (again, see Grenander and Rosenblatt (1953) and Parzen (1957), as well as Dahlhaus (1985)), here we provide a fixed-bandwidth ratio treatment. We discuss the estimation of the limit distribution, and how this can be utilized to construct spectral confidence bands.

The limit distributions $S_\theta(b)$ do not differ tremendously from the frequency zero case, but there are a few alterations from the previous distribution theory (aside from the impact of kinks in the taper) given in McElroy and Politis (2012). For all frequencies except 0 and π, the estimates converge to the sum of two independent copies of the limit in the HAC case (frequency zero); in the case of a short memory process, this result can also be found in HV, but our results also cover long memory and negative memory processes. Moreover, we focus our treatment on spectral estimates that are centered by the sample mean (so we do not consider more complicated mean regression functions), which only affects the asymptotic distribution at frequency zero. Without the centering, the limit random variable at frequency zero is a quadratic functional of Fraction Brownian Motion (FBM), instead of Fractional Brownian Bridge (FBB)—see the discussion in HV and McElroy and Politis (2012). For the numerical studies, we have simulated the limit distributions for the internal frequencies (i.e., the interval $(0, \pi)$) and the boundary frequencies (i.e., 0 and π) using FBM (because the FBB case is already addressed in McElroy and Politis (2012)), and tabulated the results by taper, bandwidth fraction b, and memory parameter in the expanded version of this paper McElroy and Politis (2014), available at http://escholarship.org/uc/item/6164c110.

Previous work (McElroy and Politis, 2011) shows the impact of memory on critical values, and that the effect is more pronounced with small b. We repeat some of this material for the spectral case, discussing the critical values as a function of b for various memory parameters. When memory is absent from all frequencies of interest, we can construct confidence intervals using the short memory critical values, but otherwise some estimate of the memory parameter must be supplied to the quantile function. In our applications we propose a simplistic nonparametric estimate of the memory parameter, as a function of frequency, and utilize a plug-in approach to inference. Our simulation studies illustrate how size is contingent on taper, bandwidth, and sample size, presuming that the memory parameter is known. These points and the general methodology are demonstrated on one construction and one retail series.

The paper is organized as follows. In Section 2 we provide a discussion of cyclical long memory, which sets the general framework for most of the paper. Then Section 3 provides the asymptotic theory for fixed-bandwidth fraction estimation of the spectral density and the spectral distribution function. In Section 4 a treatment of higher-order accuracy, with an application of the method of Laplace inversion is provided. Section 5 contains a description of our methods of simulation for critical values, as well as the performance on finite samples from simulation. The full methodology is demonstrated on two economic time series in Section 6, and Section 7 concludes. Proofs, as well as extended discussions of numerical and empirical results, are available in the fuller version of the paper available online (McElroy and Politis, 2014).

2. Cyclical long memory and data assumptions

From now on, let $\{Y_t\}$ be a constant mean stationary time series with finite variance, such that $\{\gamma_h\}$ is the autocovariance

function (acf). We define cyclical long memory in analogy with conventional long memory, such that the definition agrees with the implicit definition in seasonal fractionally integrated processes (Gray et al., 1989) and Gegenbauer processes (Woodward et al., 1998). When the acf is absolutely summable, the spectral density $f(\theta) = \sum_h \gamma_h \cos(\theta h)$ is well-defined, but here we consider the case where the spectral density has long memory poles. On the other hand, if the spectral density has a zero, this corresponds to cyclical negative memory (McElroy and Politis, 2011). We say that the time series has cyclical memory at frequency $\theta \in [0, \pi]$ if

$$\sum_{|h| \le n} \gamma_h \cos(\theta h) = L_\theta(n)\, n^{\beta_\theta}, \tag{2}$$

where L_θ is a slowly-varying function at infinity (let $\mathbb{L}$ denote the set of such functions), with a limit of $C_\theta \in [0, \infty]$. Also the memory parameter is β_θ, a number in $(-1, 1)$. The case that $\beta_0 = -1$ was explored in McElroy and Politis (2011), and it produces somewhat non-standard asymptotic results for the sample mean; we ignore this case in this paper.

Definition 1. A weakly stationary time series with spectral density f has *cyclical memory* at frequency $\theta \in [0, \pi]$ if (2) holds. This property is denoted by CM(β, θ).

Note that CM(0, θ) denotes short memory at frequency θ, i.e., $0 < f(\theta) < \infty$. More generally, the definition of cyclical memory indicates that $f(\theta)$ equals 0, ∞, or C_θ depending on whether β_θ is negative, positive, or equal to zero, and these cases correspond to negative cyclical memory, long cyclical memory, and short cyclical memory respectively (for short cyclical memory, we also impose that C_θ is a nonzero finite constant).

This is a time domain formulation of the basic concept. The following proposition relates it to a frequency domain formulation, which some readers may find more intuitive. When a zero or pole occurs at a nonzero frequency, it must be present at the negative of that frequency as well, because the spectral density is an even function on $[-\pi, \pi]$. When the zero or pole occurs at frequency zero, the spectral density might be written as $f(\lambda) = |\lambda|^\alpha g(\lambda) L(|\lambda|^{-1})$ for $\alpha \in (-1, 1)$, g a positive, even, and bounded function, and $L \in \mathbb{L}$. But if the zero/pole occurs at a nonzero frequency θ, we can generally write the spectral density as

$$f(\lambda) = |\lambda - \theta|^\alpha |\lambda + \theta|^\alpha g(\lambda)\, L(|\lambda - \theta|^{-1})\, L(|\lambda + \theta|^{-1}). \tag{3}$$

This form only treats one zero/pole frequency θ, but the following result can be easily generalized to spectra with multiple distinct zeros and/or poles.

Proposition 1. *Suppose $\{Y_t\}$ is a stationary time series with spectral density with a zero/pole of order α at frequency θ. If $\theta = 0$ and $f(\lambda) = |\lambda|^\alpha g(\lambda) L(|\lambda|^{-1})$, then the process is CM($-\alpha$, 0) and CM(0, ω) for all $\omega \neq 0$ (i.e., it has short memory at all nonzero frequencies). If $\theta > 0$ and the spectrum is given by (3), then the process is CM($-\alpha$, θ) and CM(0, ω) for $\omega \neq \theta$.*

So the processes discussed in Proposition 1 have zeros/poles of diverse orders at differing frequencies, and this in turn is connected to rates of convergence of the partial sums of autocovariances weighted by cosines. Consider the following class of spectral densities, where there are J zeros/poles at nonzero frequencies θ_j (not including the conjugate zeros/poles $-\theta_j$) of order α_j, and accompanying slowly varying functions L_j. A process with such a spectral density belongs to the class $\bigcap_{j=1}^J$ CM($-\alpha_j, \theta_j$), noting that CM($-\alpha, \theta$) = CM($-\alpha, -\theta$).

In order to formulate the asymptotic results of this paper, we must make some additional assumptions about the observed stochastic process. We will consider the same set of assumptions discussed in McElroy and Politis (2011), namely that the data process is either linear, or can be written as a function of a Gaussian process, or satisfies certain higher order cumulant conditions. The kth order cumulant of $\{Y_t\}$ is defined by

$$c_k(u_1, u_2, \ldots, u_{k-1}) = \text{cum}\left(Y_{t+u_1}, Y_{t+u_2}, \ldots, Y_{t+u_{k-1}}, Y_t\right)$$

for any t and integers $u_1, \ldots, u_{k-1}$, where $k \ge 1$ (cf. Taniguchi and Kakizawa (2000)). Letting u denote the $k-1$ vector of indices, we will write $c_k(u)$ for short. Also let $\|\cdot\|$ denote the sup-norm of a vector, so that $\sum_{\|u\| < n} c_k(u)$ is a short-hand for summing the cumulant over all indices such that $|u_j| < n$ for each j. We also require the concept of Hermite rank (Taqqu, 1975): if $g \in \mathbb{L}^2(\mathbb{R}, e^{-x^2/2})$, then it can be expanded in terms of the Hermite polynomials H_k, with coefficients $\langle g, H_k \rangle$ (the bracket denotes the inner product of the Hilbert Space) for $k \ge 0$. The Hermite rank is the index of the first nonzero coefficient.

In addition to supposing that the process is CM($\beta_{\theta_j}, \theta_j$) for a collection of frequencies $\theta_j \in [0, \pi], j = 1, \ldots, J$, we also consider the following assumptions:

- *Process* P1. $\{Y_t\}$ is a linear process: $Y_t = \sum_j \psi_j \epsilon_{t-j}$ with $\{\psi_j\}$ square summable and $\{\epsilon_t\}$ *iid* with finite variance.
- *Process* P2. $Y_t = g(X_t)$ for each t, where g is a function in $\mathbb{L}^2(\mathbb{R}, e^{-x^2/2})$ of Hermite rank τ, and $\{X_t\}$ is a Gaussian process with autocovariance function r_k. If $\beta_{\theta_j} > 0$, also assume that $(1 - \beta_{\theta_j})\tau < 1$ for each j.
- *Process* P3. $\{Y_t\}$ is a strictly stationary process whose kth order cumulants exist and are summable over its k indices, for all $k \ge 1$. Moreover, when $\beta_{\theta_j} < 0$ we also assume that $\sum_{\|u\| < n} c_k(u) = O(n^{\beta_{\theta_j}})$ for each j.

See the discussion in McElroy and Politis (2011) for why a moment-plus-mixing condition is not viable. Each of the assumptions P1, P2, or P3 is sufficient to establish a limit theorem for the Discrete Fourier Transforms of the data, as shown below. These process assumptions are typically unverifiable from the observed data, and should be viewed as working assumptions.

3. Asymptotic theory for spectrum estimation

The theory developed here is similar to that of HV, but is extended to processes with cyclical memory, similarly to how McElroy and Politis (2012) extended the HAC theory to long-range dependent processes. First we establish a joint convergence theorem for normalized Discrete Fourier Transforms (DFTs), which is a result of independent interest. Second, we apply this result to the analysis of taper-smoothed estimates of the spectral density. Third, we address the estimation of the spectral distribution function in the case of a bounded positive spectral density.

3.1. Theory for DFTs

Let $\{Y_t\}$ be a mean μ stationary time series with acf $\{\gamma_h\}$, as described in Section 2. We suppose that a sample of size N is available: $Y_1, Y_2, \ldots, Y_N$, and the sample autocovariances are computed via

$$\widehat{\gamma}_h = \frac{1}{N} \sum_{t=1}^{N-h} (Y_t - \overline{Y})\,(Y_{t+h} - \overline{Y})$$

for $h = 0, 1, 2, \ldots$, and $\overline{Y} = n^{-1} \sum_{t=1}^n Y_t$. Results can be modified easily if we do not demean and assume $\mu = 0$ (as discussed in HV as well), but our main exposition assumes centering of estimates by the sample mean for simplicity of presentation. The DFT of the sample is $\sum_{t=1}^N (Y_t - \overline{Y}) e^{-i\theta t}$, which has real and imaginary parts given by cosine and sine summations, respectively. These trigonometric partial sums are the key aspect in the asymptotic analysis

of the spectral density estimates of this paper. We introduce the weighted-sum notation as follows:

$$S_N(g) = \sum_{t=1}^{N} Y_t g_t$$

for a sequence $\{g_t\}$. Then the DFT equals $S_N(c(\theta)) + iS_N(s(\theta))$ for $c(\theta) = \cos(\theta\cdot)$ and $s(\theta) = \sin(\theta\cdot)$. The rate of growth of $S_N(c(\theta))$ and $S_N(s(\theta))$ will depend upon θ, because if there is a zero or pole at frequency θ the growth rate is affected by long-range dependence. Ultimately, we wish to prove joint functional limit theorems for the processes $r \mapsto \{S_{[rN]}(c(\theta)), S_{[rN]}(s(\theta))\}$, jointly over a finite collection of frequencies θ. Here the square bracket refers to the greatest integer function.

The key quantities that determine the growth rates of the real and imaginary parts of the DFT are the respective variances:

$$V_N^+(\theta) = \mathrm{Var} S_N(c(\theta)) \qquad V_N^-(\theta) = \mathrm{Var} S_N(s(\theta)).$$

When $\theta \neq 0, \pi$, we let $V_N(\theta) = (V_N^+(\theta) + V_N^-(\theta))/2$, but for $\theta = 0, \pi$ we set $V_N(\theta) = V_N^+(\theta)$ instead. Then with $W_N(\theta) = \sum_{|h|\leq N} \gamma_h \cos(\theta h)$, we have the following identity:

$$V_N(\theta) = \frac{1 + 1_{\{\theta_j = 0, \pi\}}}{2} \sum_{k=0}^{N-1} W_k(\theta). \tag{4}$$

This follows by recognizing that

$$V_N(\theta) = \frac{1 + 1_{\{\theta_j = 0, \pi\}}}{2} \sum_{j,k=1}^{N} \cos(\theta(j-k))\gamma_{j-k}, \tag{5}$$

and that the latter expression in (5) can be re-expressed, using summation by parts, into (4). Noting that the definition of $W_N(\theta)$ together with the CM(β_θ,θ) assumption yields an asymptotic growth rate of $L_\theta(N)\, N^{\beta_\theta}$, we can apply (4) and Proposition 1 of McElroy and Politis (2011) to the autocovariance sequence $\{\gamma_h \cos(\theta h)\}$ for any θ to obtain

$$V_N(\theta) \sim \frac{L_\theta(N)\, N^{\beta_\theta + 1}}{2(\beta_\theta + 1)}. \tag{6}$$

In the case of short memory, where $\beta_\theta = 0$ and L_θ tends to a nonzero constant C_θ, (6) becomes $V_N(\theta) \sim N\, C(\theta)$ and $C(\theta)$ equals one half the spectral density. In all cases of cyclical memory, the square root of $V_N(\theta)$ will be the appropriate normalizing rate for the DFT sums, as shown below.

Let us consider a finite collection of J distinct frequencies $\Theta = \{\theta_j\}_{j=1}^J$ in $[0, \pi]$, where the data process is CM(β_{θ_j}, θ_j) for each j. Define vector-valued stochastic processes as follows:

$$S_{[rN]}(c(\theta)) = \{S_{[rN]}(c(\theta_j))\}_{j=1}^J \qquad S_{[rN]}(s(\theta)) = \{S_{[rN]}(s(\theta_j))\}_{j=1}^J,$$

where $r \in [0, 1]$. Joint functional limit theorems for $S_{[\cdot N]}(c(\theta))$ and $S_{[\cdot N]}(s(\theta))$ normalized each by $V_N^{1/2}(\theta)$ form the key foundation for the asymptotic theory for the tapered-estimates of the spectral density, defined in the next subsection. The limit stochastic processes are $B_{+,\theta}(\cdot) = \{B_{+,\theta_j}(\cdot)\}_{j=1}^J$ and $B_{-,\theta}(\cdot) = \{B_{-,\theta_j}(\cdot)\}_{j=1}^J$, all of which are independent of each other, and all of which are Fractional Brownian Motions (FBMs) of parameter β_{θ_j}, except $B_{-,\theta}(\cdot)$ at $\theta = 0, \pi$, which is the zero process.

As discussed in McElroy and Politis (2011), it is more convenient for us to formulate the results in the space $C[0, 1]$ of continuous functions, rather than the Skorohod space. Therefore we will consider a linearly-interpolated version $\xi_{[\cdot N]}(g)$ of $S_{[\cdot N]}(g)$, defined via $\xi_{[rN]}(g) = S_{[rN]}(g) + (rN - [rN])Y_{[rN]+1}$. This affects the mean-centering slightly, though the asymptotic impact is negligible. Define the functions $c_N(\theta) = \sum_{t=1}^N \cos(\theta t)$ and $s_N(\theta) = \sum_{t=1}^N \sin(\theta t)$, which mean center $S_N(c(\theta))$ and $S_N(s(\theta))$ respectively. The mean-centering functions for $\xi_{[rN]}(c(\theta_j))$ and $\xi_{[rN]}(s(\theta_j))$ are given by

$$\widetilde{\mu}_r(c(\theta_j)) = \mu\, c_{[rN]}(\theta_j) + \mu\, (rN - [rN])\, \cos(\theta_j([\cdot N] + 1))$$
$$\widetilde{\mu}_r(s(\theta_j)) = \mu\, s_{[rN]}(\theta_j) + \mu\, (rN - [rN])\, \sin(\theta_j([\cdot N] + 1)),$$

respectively.

Theorem 1. *Let $\{Y_t\}$ be covariance stationary with mean μ and acf $\{\gamma_h\}$, such that the process is CM(β_{θ_j}, θ_j) for a collection of frequencies $\theta_j \in [0, \pi]$, $j = 1, \ldots, J$. Letting $\kappa = \max_{1\leq j\leq J} 2 \wedge [2/(1 + \beta_{\theta_j})]$, suppose that $\mathbb{E}[|Y_t|^{\kappa+\delta}] < \infty$ for some $\delta > 0$, and also assume that $\mathbb{E}[|S_n(c(\theta_j)) - c_n(\theta_j)|^{\kappa+\delta}] = O(V_n^{(\kappa+\delta)/2}(\theta_j))$ and $\mathbb{E}[|S_n(s(\theta_j)) - s_n(\theta_j)|^{\kappa+\delta}] = O(V_n^{(\kappa+\delta)/2}(\theta_j))$ hold. Suppose condition* P1, P2, *or* P3 *holds, and that in the case of a* P2 *process with at least one $\beta_{\theta_j} > 0$, the Hermite rank is unity. Then the following weak convergence holds in the space $\mathcal{C}([0, 1], \mathbb{R}^{2J})$:*

$$\Big\{V_N^{-1/2}(\theta_j)\, \big(\xi_{[\cdot N]}(c(\theta_j)) - \widetilde{\mu}_\cdot(c(\theta_j))\big)\,,$$
$$V_N^{-1/2}(\theta_j)\, \big(\xi_{[\cdot N]}(s(\theta_j)) - \widetilde{\mu}_\cdot(s(\theta_j))\big)\Big\}_{j=1}^J$$
$$\stackrel{\mathcal{L}}{\Longrightarrow} \{B_{+,\theta_j}, B_{-,\theta_j}\}_{j=1}^J. \tag{7}$$

Remark 1. By 1.342.2 of Gradshteyn and Ryzhik (1994), $c_N(\theta)$ equals N if θ is an integer multiple of 2π, and otherwise equals

$$\frac{1}{2}\left[\frac{\sin((N + 1/2)\theta)}{\sin(\theta/2)} - 1\right].$$

Also by 1.342.1 of Gradshteyn and Ryzhik, $s_N(\theta)$ equals 0 if θ is an integer multiple of π, and otherwise equals

$$\sin[(N + 1)\theta/2]\, \sin[N\theta/2]\, \csc[\theta/2].$$

Hence the centering for the sine partial sum is asymptotically irrelevant, as is the centering for the cosine partial sum unless $\theta = 0$.

Theorem 1 provides the assumed conditions (4), (5), (6), and (7) of HV, and also provides a generalization of the short memory situation. We next discuss its application to spectral density estimation.

3.2. Asymptotic theory for spectral density estimation

Now in order to apply (7) to spectral estimation, it is necessary to extend the FBMs discussed above to Fractional Brownian Bridges (FBBs) as in HV, defined as follows:

$$\widetilde{B}_{\pm,\theta}(r) = B_{\pm,\theta}(r) - 1_{\{\theta=0\}} \int_0^r x'(t)\, dt \left[\int_0^1 x(t)x'(t)\, dt\right]^{-1} \times \int_0^1 x(t)\, dB_{\pm,\theta}(t).$$

Here x is a deterministic vector process with each component $x^j \in C[0, 1]$, and corresponds to regression effects in the data process; see Phillips (1998) for a more detailed exposition. That is, when the mean of the process $\{Y_t\}$ is non-constant, and perhaps is parametrized by regression functions such that the demeaned $\{Y_t\}$ is mean zero and stationary, then our partial sums and DFT statistics should be constructed from variables Y_t centered by estimates of these mean effects. In this paper, we focus on the simple case that $x(t) \equiv 1$, corresponding to centering by the sample mean (the ordinary least squares estimate of a constant

mean); see the Appendix of McElroy and Politis (2014) for a partial elaboration of the more general case. Note that this centering has no impact except at frequency zero, which follows from Remark 1 above, which shows that only the real part of the DFT (i.e., the cosine partial sum) at frequency zero needs to be mean-centered. In the case that the true mean is zero and this assumption is utilized in our statistics, then $x(t) \equiv 0$ and $\widetilde{B}_{\pm,\theta} = B_{\pm,\theta}$, the FBM. But when we center by the sample mean, it follows that $\widetilde{B}_{\pm,0}(r) = B_{\pm,0}(r) - rB_{\pm,0}(1)$, a FBB.

We now suppose that an estimate of the spectrum is computed via (1) using autocovariance estimates centered by the sample mean (or without centering in the special case that the mean is known to be zero), as described above. The taper (or kernel) Λ comes from a wide family that encompasses flat-top tapers (Politis, 2001), the Bartlett taper, as well as other tapers considered in Kiefer and Vogelsang (2005) and HV:

$$\{\Lambda \text{ is even with support on } [-1,1] \text{ such that } \Lambda(x) \text{ is constant for } |x| \le c, \text{ for some } c \in [0,1); \text{ also, } \Lambda \text{ is twice continuously differentiable on } (c,1)\}. \quad (8)$$

A derivative of Λ from the left (with respect to x) is denoted by $\dot{\Lambda}^-$, whereas from the right is $\dot{\Lambda}^+$; the second derivative is $\ddot{\Lambda}$. Note that we allow for Λ to have a jump discontinuity at c; for example, our results apply to the truncation taper given by the indicator on the interval $[-c, c]$. Our main result, which is stated next, follows from Theorem 1 and an analysis of the spectral estimator, expanding on the analysis of HV.

Theorem 2. *Let $\{Y_t\}$ be covariance stationary with mean μ and acf $\{\gamma_h\}$, such that the process is CM(β_{θ_j}, θ_j) for a collection of frequencies $\theta_j \in [0,\pi]$, $j = 1,\ldots,J$. Letting $\kappa = \max_{1\le j\le J} 2 \wedge [2/(1+\beta_{\theta_j})]$, suppose that $\mathbb{E}[|Y_t|^{\kappa+\delta}] < \infty$ for some $\delta > 0$, and also assume that $\mathbb{E}[|S_n(c(\theta_j)) - c_n(\theta_j)|^{\kappa+\delta}] = O(V_n^{(\kappa+\delta)/2}(\theta_j))$ and $\mathbb{E}[|S_n(s(\theta_j)) - s_n(\theta_j)|^{\kappa+\delta}] = O(V_n^{(\kappa+\delta)/2}(\theta_j))$ hold. Suppose condition* P1, P2, *or* P3 *holds, and that in the case of a* P2 *process with at least one $\beta_{\theta_j} > 0$, the Hermite rank is unity. Also suppose that either the sample autocovariances are centered by the sample mean, or they are not centered and that $\mu = 0$. For tapers defined via (8), as $N \to \infty$ we have*

$$\begin{aligned}\frac{N\widehat{f}(\theta_j)}{V_N(\theta_j)} \overset{\mathcal{L}}{\Longrightarrow} & -\frac{1}{b^2}\int\int_{cb<|r-s|<b} \ddot{\Lambda}\left(\frac{r-s}{b}\right) \\ & \times \left(\widetilde{B}_{+,\theta_j}(r)\widetilde{B}_{+,\theta_j}(s) + \widetilde{B}_{-,\theta_j}(r)\widetilde{B}_{-,\theta_j}(s)\right) dr\,ds \\ & + \frac{2}{b}\dot{\Lambda}_-(1)\int_0^{1-b}\left(\widetilde{B}_{+,\theta_j}(r)\widetilde{B}_{+,\theta_j}(r+b)\right. \\ & \left. + \widetilde{B}_{-,\theta_j}(r)\widetilde{B}_{-,\theta_j}(r+b)\right) dr \\ & - \frac{2}{b}\dot{\Lambda}_+(c)\int_0^{1-bc}\left(\widetilde{B}_{+,\theta_j}(r)\widetilde{B}_{+,\theta_j}(r+bc)\right. \\ & \left. + \widetilde{B}_{-,\theta_j}(r)\widetilde{B}_{-,\theta_j}(r+bc)\right) dr \\ & + \frac{2}{b}\int_{1-b}^{1-bc}\dot{\Lambda}\left(\frac{1-r}{b}\right)\left(\widetilde{B}_{+,\theta_j}(r)\widetilde{B}_{+,\theta_j}(1)\right. \\ & \left. + \widetilde{B}_{-,\theta_j}(r)\widetilde{B}_{-,\theta_j}(1)\right) dr \\ & + \Lambda(0)\left(\widetilde{B}^2_{+,\theta_j}(1) + \widetilde{B}^2_{-,\theta_j}(1)\right),\end{aligned}$$

jointly in θ_j for $j = 1,2,\ldots,J$. In the case that there is a jump discontinuity in Λ at c, we must replace the third summand in the limit distribution by

$$2\left(\Lambda^+(c) - \Lambda^-(c)\right) \times \left(\widetilde{B}_{+,\theta_j}(1-bc)\widetilde{B}_{+,\theta_j}(1) + \widetilde{B}_{-,\theta_j}(1-bc)\widetilde{B}_{-,\theta_j}(1)\right).$$

Allowing for jump discontinuities in the taper extends the results of HV; in particular, utilizing the taper $\Lambda = 1_{[-1,1]}$ and $b = 1$ corresponds to the periodogram, and the limit result is $\widetilde{B}^2_{+,\theta_j}(1) + \widetilde{B}^2_{-,\theta_j}(1)$, or a χ^2 on two degrees of freedom (i.e., the classical result). More generally, the result applies for flat-top tapers (e.g., the trapezoidal), tapers with kinks (e.g., the Bartlett), and smooth tapers (e.g., the Parzen). The theorem describes the limit behavior of the spectral density estimate in the case that cyclical memory is present, considering a finite collection of frequencies. If these frequencies happen to correspond to short memory dynamics, then the spectral density is finite and nonzero. Letting $\tau_\theta = \frac{1+1_{\{\theta=0,\pi\}}}{2}$, from (4) we have

$$V_N(\theta) \sim N\,\tau_\theta\, f(\theta), \quad (9)$$

so that the convergence of Theorem 2 in the case of short memory may be summarized as

$$\widehat{f}(\theta) \overset{\mathcal{L}}{\Longrightarrow} \tau_\theta\, f(\theta)\, S_\theta(b),$$

where we denote the limit random variable on the right hand side of the convergence in Theorem 2 via $S_\theta(b)$. A numerical description of this distribution is given in HV. A technical description can be given through the moment generating function, or Laplace Transform (LT) of $S_\theta(b)$, as in McElroy and Politis (2009); this is developed in Section 4 below. Tables of quantiles can be given over a grid of b values, depending on the three frequency cases (i.e., $\theta = 0$, $\theta = \pi$, or $\theta \in (0,\pi)$) and the taper; see McElroy and Politis (2014) for details (http://escholarship.org/uc/item/6164c110).

In the case of cyclical long memory or negative memory the true spectrum $f(\theta)$ is either equal to ∞ or zero, and inference is problematic. For the purpose of constructing a confidence interval, we propose the quantity $f_N(\theta) = V_N(\theta)/(N\tau_\theta)$ as the "parameter" of interest, although clearly this is a moving target; only in the case of short memory can we conceptually replace $f_N(\theta)$ by $f(\theta)$, via (9). However, whatever the degree of cyclical memory, we can conduct inference for $f_N(\theta)$ as follows. Denote the quantile function of $S_\theta(b)$ by $Q_\theta(\cdot)$. If we wish to consider a single frequency, the confidence interval for $f_N(\theta)$ with asymptotic coverage $1-\alpha$ is

$$\left[\frac{\widehat{f}(\theta)}{\tau_\theta\, Q_\theta(1-\alpha/2)}, \frac{\widehat{f}(\theta)}{\tau_\theta\, Q_\theta(\alpha/2)}\right], \quad (10)$$

which follows from $\mathbb{P}\left[Q_\theta(\alpha/2) \le \frac{\widehat{f}(\theta)}{\tau_\theta f_N(\theta)} \le Q_\theta(1-\alpha/2)\right] \to 1-\alpha$. Alternatively, a simultaneous confidence interval can be constructed by considering the maximum and minimum of $S_\theta(b)$ over the pertinent frequencies. Let $\overline{S}(b) = \max_{1\le j\le J} S_{\theta_j}(b)/\tau_{\theta_j}$ and $\underline{S}(b) = \min_{1\le j\le J} S_{\theta_j}(b)/\tau_{\theta_j}$, which have distributions easily computable from the marginals due to independence (they are also identically distributed for $\theta_j \in (0,\pi)$). (Note that our notation assumes that the same bandwidth fraction b is used for all frequencies, although this need not be the case in practice.) The corresponding quantile functions will be denoted by $\overline{Q}$ and $\underline{Q}$ for the maximum and minimum respectively. Let J denote a finite index set, and consider a set of frequencies θ_j with $1 \le j \le J$. For positive real numbers ℓ, u, we have

$$\begin{aligned}\mathbb{P}\left[\frac{\widehat{f}(\theta_j)}{\tau_{\theta_j} u} \le f_N(\theta_j) \le \frac{\widehat{f}(\theta_j)}{\tau_{\theta_j}\,\ell}\ \forall j\right] &= \mathbb{P}\left[\ell \le \frac{\widehat{f}(\theta_j)}{\tau_{\theta_j} f_N(\theta_j)} \le u\ \forall j\right] \\ \to \mathbb{P}\left[\ell \le \underline{S}(b) \le \overline{S}(b) \le u\right] &= 1 - \mathbb{P}\left[\underline{S}(b) \le \ell\right] - \mathbb{P}\left[\overline{S}(b) \ge u\right].\end{aligned}$$

The last equality follows from the observation that – when $\ell < u$ – the event $\{\underline{S} \le \ell\}$ is mutually exclusive with the event $\{\overline{S} \ge u\}$. This probability is approximately $1-\alpha$ if ℓ, u correspond to the appropriate critical values; splitting the quantity α evenly amounts to

$$\ell = \underline{Q}(\alpha/2) \qquad u = \overline{Q}(1-\alpha/2). \quad (11)$$

This provides the construction of a simultaneous confidence interval.

3.3. Asymptotic theory for spectral distribution estimation

The estimation of spectral content can be extended to the spectral distribution function $F(\theta) = (2\pi)^{-1}\int_{-\pi}^{\theta} f(\lambda)\,d\lambda$, and because of the smoothing of the spectral density accomplished by integration, the behavior of statistical estimates is easier to describe. In this subsection we assume that the spectral density has short memory, and hence $0 < f(\lambda) < \infty$ for all $\lambda \in [-\pi, \pi]$. We make this assumption so that the rate of convergence of spectral estimates is the same at all frequencies. Indeed, the classical limit result of Dahlhaus (1985) cannot hold for processes with long memory poles such that $\beta > 1/2$, because the limiting variance (see below) depends on the integral of the squared spectral density.

Because the spectral density is even, it suffices to study $G(\theta) = (2\pi)^{-1}\int_0^{\theta} f(\lambda)\,d\lambda$, and its corresponding estimator $\widehat{G}(\theta) = (2\pi)^{-1}\int_0^{\theta}\widehat{f}(\lambda)\,d\lambda$. Very general results for functionals of the periodogram, under general data process conditions, were obtained by Dahlhaus (1985); also see the literature cited in that paper for a history of efforts. Whereas Dahlhaus (1985) utilizes a data taper, here we utilize a covariance taper – in keeping with the previous subsection on spectral density estimation – as other literature has also done (e.g., Priestley (1981)). The novelty of this subsection lies chiefly in adopting a fixed bandwidth ratio framework, and somewhat unsurprisingly the same limit distribution and functional limit theorem is obtained as in Dahlhaus (1985); in particular, neither bandwidth fraction b nor taper plays any role in the asymptotic distribution.

Utilizing the definition of the spectral density estimator, we at once obtain

$$\widehat{G}(\theta) = \widehat{\gamma}_0\frac{\theta}{2\pi} + 2\sum_{h=1}^{N-1}\Lambda(h/bN)\,\widehat{\gamma}_h\,\frac{\sin[\theta h]}{2\pi h} = \frac{1}{2\pi}\int_{-\pi}^{\pi} I(\lambda)\sum_{|h|<N}\Lambda(h/bN)\,\frac{\sin[\theta h]}{2\pi h}\,e^{i\lambda h}\,d\lambda,$$

where we interpret $\sin[\theta h]/h$ to be the value θ whenever $h = 0$. Here $I(\lambda)$ is the periodogram, defined to be N^{-1} times the magnitude squared of the DFT:

$$I(\lambda) = N^{-1}\left|\sum_{t=1}^{N}(Y_t - \overline{Y})e^{-i\lambda t}\right|^2.$$

Let $g_\theta(\lambda) = \Lambda(0)\sum_{h\in\mathbb{Z}}\frac{\sin[\theta h]}{2\pi h}e^{i\lambda h}$, which is the pointwise limit of

$$g_{N,\theta}(\lambda) = \sum_{|h|<N}\Lambda(h/bN)\,\frac{\sin[\theta h]}{2\pi h}\,e^{i\lambda h}.$$

Because of symmetry, g_θ is always real, and so the complex exponential can be replaced by a cosine in its definition. We claim that this pointwise limit can be taken in the definition of $\widehat{G}(\theta)$. Note that $g_\theta(\lambda) = 2^{-1}\,1_{[-\theta,\theta]}(\lambda)$, the sinc function. Let $\widehat{G}(\cdot)$ denote the spectral distribution function's estimate, and the limiting process $\mathcal{Z}(\cdot)$ is defined as a mean zero Gaussian process with covariance kernel

$$K(\theta,\omega) = \mathrm{Cov}\,(\mathcal{Z}(\theta),\mathcal{Z}(\omega)) = \pi^{-1}\int_{-\pi}^{\pi} g_\theta(\lambda)g_\omega(\lambda)f^2(\lambda)\,d\lambda = \frac{1}{2\pi}\int_0^{\theta\wedge\omega} f^2(\lambda)\,d\lambda. \tag{12}$$

This kernel is simpler than the one found in Dahlhaus (1985), because we will assume that fourth order cumulants are zero (this could be relaxed, but then a different approach to the estimation of limit quantiles in Theorem 3 would be needed). The kernel actually corresponds to the covariance kernel of a heteroscedastic Brownian Motion (see below).

We focus on $G(\theta)$ rather than $F(\theta)$, because if we are interested in $F(\theta)$ for $\theta < 0$, this is equal to $G(\pi) - G(-\theta)$ by symmetry. So the following functional limit theorem can be stated; like Theorem 4.1 of Dahlhaus (1985) we require eighth order moments. (Using the less restrictive tightness criterion described in Karatzas and Shreve (1991), one could relax the requirement to $4+\delta$ moments, for some $\delta > 0$, but then other conditions – that are harder to verify – would have to be added to compensate.)

Theorem 3. *Let $\{Y_t\}$ be covariance stationary with mean μ and acf $\{\gamma_h\}$, such that the process has short memory, satisfies $\mathbb{E}[Y_t{}^8] < \infty$, and such that condition P1, P2, or P3 holds. Also suppose that the fourth order cumulants are zero. If the taper satisfies (8), then as $N \to \infty$ we have*

$$\sqrt{N}\,\left(\widehat{G}(\cdot) - G(\cdot)\right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{Z}(\cdot)$$

in the space $\mathcal{C}([0,\pi],\mathbb{R})$, where the process $\mathcal{Z}$ is mean zero Gaussian with covariance kernel (12).

It is interesting that the taper is irrelevant to the asymptotic distribution—this is essentially because the integration involved in the definition of the spectral distribution makes the tapering in the spectral density estimation obsolete. However, the taper and the bandwidth have a substantial impact on the qualitative features of the estimate (see Section 5). The degree of correlation between differing values of the spectral distribution estimator depends chiefly on the smaller frequency, as indicated by (12); variance is increasing in frequency, unto the maximum value $G(\pi) = \gamma_0/2$.

As an application of Theorem 3, we can construct uniform confidence bands about the spectral distribution function. This is in contrast to the application discussed in Section 3.2, where simultaneous confidence intervals were constructed for a finite number of frequencies. For real numbers ℓ, u we have the confidence band $[\widehat{G}(\theta) - u/\sqrt{N}, \widehat{G}(\theta) - \ell/\sqrt{N}]$ – as a function of $\theta \in [0,\pi]$ – yielding coverage as follows:

$$\begin{aligned}&\mathbb{P}\left[\widehat{G}(\theta) - u/\sqrt{N} \le G(\theta) \le \widehat{G}(\theta) - \ell/\sqrt{N}, \forall\theta\in[0,\pi]\right]\\ &= 1 - \mathbb{P}\left[\sup_{\theta\in[0,\pi]}\sqrt{N}\,(\widehat{G}(\theta) - G(\theta)) \ge u\right]\\ &\quad - \mathbb{P}\left[\inf_{\theta\in[0,\pi]}\sqrt{N}\,(\widehat{G}(\theta) - G(\theta)) \le \ell\right]\\ &\to 1 - \mathbb{P}\left[\sup_{\theta\in[0,\pi]}\mathcal{Z}(\theta) \ge u\right] - \mathbb{P}\left[\inf_{\theta\in[0,\pi]}\mathcal{Z}(\theta) \le \ell\right]\end{aligned}$$

as $N \to \infty$. The random variables $\underline{Z} = \inf_{\theta\in[0,\pi]}\mathcal{Z}(\theta)$ and $\overline{Z} = \sup_{\theta\in[0,\pi]}\mathcal{Z}(\theta)$ determine the spread of the confidence band, and can be calculated via simulation when the covariance kernel is known, or is estimable. Another possibility is to estimate the limit distribution via subsampling (this might be preferable if the assumption on the fourth cumulant is not tenable), as in Politis et al. (1993).

Let the corresponding quantile functions be denoted by $\underline{R}$ and $\overline{R}$ respectively. Then the confidence band probability is approximately $1-\alpha$ if ℓ, u correspond to the appropriate critical values; splitting α evenly yields

$$\ell = \underline{R}(\alpha/2) \qquad u = \overline{R}(1-\alpha/2). \tag{13}$$

This construction differs somewhat from (11), because in that case the limit theorem was formulated as a ratio (for spectral density estimation), whereas here the limit theorem is formulated as a difference (for spectral distribution estimation). Although the limit

$\mathcal{Z}(\theta)$ does not depend on the taper, it does require a knowledge of f. In practice, one must construct an estimate of the covariance kernel (12); we next describe our procedure.

Let M denote a mesh of frequencies, providing a discretization of the Riemann integral defining $K(\theta, \theta)$. Then

$$\frac{1}{2\pi}\int_0^\theta f^2(\lambda)\,d\lambda \approx \frac{1}{2M}\sum_{t=0}^{\lfloor M\theta/\pi\rfloor} f^2(t\pi/M),$$

which is the variance of a heteroscedastic random walk. That is, suppose that $\{\epsilon_t\}$ is an independent Gaussian sequence, with each random variable having variance $f^2(t\pi/M)/(2M)$ for M fixed. Then $U_\ell = \sum_{t=1}^{\ell}\epsilon_t$ is a heteroscedastic random walk with variance approximately $K(\theta, \theta)$, where $\ell = \lfloor M\theta/\pi\rfloor$. We can easily simulate this Gaussian sequence by multiplying $f(t\pi/M)/\sqrt{2M}$ times iid normals. Moreover, the covariance function of the process $\{U_\ell\}$ is approximately that of the kernel K, because of the random walk structure.

If f is known (as in the case of hypothesis testing) then we can simulate the process $\{U_\ell\}$ and obtain an approximation to $\{\mathcal{Z}(\theta)\}$, with the association $\ell = \lfloor M\theta/\pi\rfloor$. However, in many applications f is unknown and must be estimated. One could use the tapered spectral density estimates discussed above, or the periodogram (integration over frequencies smooths it out sufficiently to provide consistency). Thus, we construct $\widehat{\epsilon}_t$ via multiplying $\widehat{f}(t\pi/M)/\sqrt{2M}$ by a standard normal, independently for each t, and construct the corresponding heteroscedastic random walk $\{\widehat{U}_\ell\}$. Here $\widehat{f}$ could be the periodogram or the same tapered spectral estimate upon which our original $\widehat{G}$ is based. Then with $\mathcal{Z}(\theta) = \widehat{U}_{\lfloor M\theta/\pi\rfloor}$, we approximate $\underline{Z}$ and $\overline{Z}$ by the minimum and maximum, respectively, over the M values $\widehat{U}_1, \ldots, \widehat{U}_M$. Repeated samples for $\{\widehat{\epsilon}_t\}$ then yield an estimate for the distribution of $\underline{Z}$ and $\overline{Z}$. Consistency of this implicit estimator $\widehat{K}$ follows from the same assumptions as used in Theorem 3. The upper quantile of $\overline{R}$ and lower quantile of $\underline{R}$ yield estimates of u and ℓ. This procedure has been implemented and tested in simulation (see Section 5 below).

Alternatively, one may be interested in testing some null hypothesis that naturally supplies f to us. For example, we may be studying the time series residuals arising from a fitted model, and seek to test whether these residuals behave as white noise. Ignoring issues of parameter estimation error, we wish to test whether $f(\lambda) \equiv \gamma_0$, and hence we can estimate the covariance kernel via

$$\widehat{K}(\theta, \omega) = \widehat{\gamma}_0^2\,\frac{\theta \wedge \omega}{2\pi}.$$

This is the kernel of a Brownian Motion process on $[0, \pi]$, scaled by $\widehat{\gamma}_0$. There exist published quantile functions for the supremum and infimum of BMs, and so the construction of ℓ, u is relatively straightforward. In this problem, the null hypothesis also dictates the form of G, i.e., $G(\theta) = \gamma_0\,\theta/(2\pi)$, so that if this particular function G fails to lie completely within the confidence bands, we have evidence to reject the null hypothesis. However, such an approach ultimately presumes a parametric specification for the original spectrum, and there are other techniques available for testing model goodness-of-fit in such a scenario. In our applications below, we focus upon nonparametric approaches to spectral estimation.

4. Higher order accuracy of the fixed bandwidth fraction

In this paper we have adopted the asymptotic perspective that bandwidth in spectral estimates is to be viewed as a fixed fraction b of the sample size. Conventional asymptotics stipulate that the bandwidth is vanishing relative to sample size, and the spectral estimates become consistent. As in the HAC literature – which examines the distribution of the self-normalized mean as $b \to 0$, and makes comparison to the conventional asymptotic normality results – we intend to examine the behavior of our limits $S_\theta(b)$ in Theorem 2 as $b \to 0$. The point of this is to show that $S_\theta(b)$ can be viewed as the classical limit distribution $S_\theta(0)$ plus other stochastic terms that are order b, b^2, and so forth. This will demonstrate a higher-order accuracy for the fixed bandwidth fraction asymptotics.

Unlike in the HAC case of a standardized sample mean statistic, where the $b = 0$ case corresponds to a Gaussian random variable, for spectral estimation the $b = 0$ case corresponds to point mass at the spectral density, i.e., $S_\theta(0) = f(\theta)$ with probability one. Therefore, expansions of the distribution of $S_\theta(b)$ as $b \to 0$ will use slightly different techniques than those employed in Sun et al. (2008). We pursue an analysis of the Laplace Transform of $S_\theta(b)$, providing a small b expansion, and relate this transform to the cumulative distribution function of $S_\theta(b)$. We utilize an expansion of the Laplace Transform in terms of functions that have known Laplace inverses; we believe this to be a novel method, potentially generalizable to other types of distribution problems.

This method will result in an expansion of the right tailed cumulative distribution function (cdf) in terms of polynomials and exponential functions, with coefficients given by polynomial functions of the cumulants. We show how to compute these cumulants directly from the tapers—although similar types of cumulant calculations have previously appeared in the HAC literature (Sun et al., 2008). However, we do not view this expansion as the most practical method for calculating the cdf; in practice, one wants the quantiles of the limit distribution, and these can be obtained via simulation (Section 5).

Fixing θ so that we can drop the subscript, the distribution of $S(b)$ is characterized by its Laplace Transform (LT). From Tziritas (1987), the LT of a Gaussian quadratic form $\langle Z, Z\rangle_T$ – for a Gaussian process Z with covariance kernel K, and a quadratic form $\langle\cdot,\cdot\rangle_T$ with operator T – is given by

$$\mathbb{E}\exp\{-s\langle Z, Z\rangle_T\} = \exp\left\{\sum_{j=1}^{\infty}(-1)^j\,\frac{\kappa_j}{j!}s^j\right\},$$

where κ_j is the jth cumulant of $S(b)$, and has the formula

$$\kappa_j = 2^{j-1}\,(j-1)!\,\mathrm{tr}[(KT)^j]. \tag{14}$$

Also see the discussion in McElroy and Politis (2012). Briefly, the Gaussian process Z is defined on the space of real-valued function of domain $[0, 1]$, such that the action of an operator A on any element x of this space is given by $(Ax)(s) = \int_0^1 A(s, t)x(t)dt$. In Eq. (14), both K and T are operators, and their composition has action on an element x given by

$$(KTx)(s) = \int_0^1\int_0^1 K(s, u)T(u, t)x(t)dtdu.$$

Also, tr denotes the trace of an operator, i.e., $\mathrm{tr}(A) = \int_0^1 A(s, s)\,ds$.

The limit distribution $S(b)$ in Theorem 2 is the sum of two such independent and identically distributed Gaussian quadratic forms (just one copy if $\theta = 0, \pi$), because it can be written as the sum of two random variables of the type

$$\int_0^1\int_0^1 T(r, s)\widetilde{B}(r)\widetilde{B}(s)\,drds,$$

where $T(r, s)$ is equal to $-b^{-2}\ddot{\Lambda}((r-s)/b)$ plus secondary terms involving the Dirac delta function. Because the Gaussian processes $\widetilde{B}$ are FBBs, the covariance kernel K is that of FBB (Samorodnitsky and Taqqu, 1994). Trivially, the LT of the sum of two iid random variables is the square of their common LT, which amounts to a doubling of each cumulant. In the following treatment, we provide an expansion for the cdf in terms of cumulants; these are given

by doubling the formula for κ_j in (14) when $\theta \neq 0, \pi$, but at frequency zero or π we just take the formula (14) directly. Since the trace of powers of KT is not convenient to calculate, we provide a feasible approximation to the κ_j in McElroy and Politis (2014); see http://escholarship.org/uc/item/6164c110.

The right-tailed cdf of $\langle Z, Z\rangle_T$ will be denoted by $\overline{F}$, and its pdf by p. The LT of a function ϕ (of non-negative support) is denoted by $\mathcal{L}_\phi$, where

$$\mathcal{L}_\phi(s) = \int_0^\infty \phi(x) e^{-sx}\, dx.$$

Then $\mathcal{L}_{\overline{F}}(s) = s^{-1}(1 - \mathcal{L}_p(s))$ using integration by parts, and $\mathcal{L}_p(s) = \mathbb{E}\exp\{-s\langle Z, Z\rangle_T\}$. Next, letting $\sum_{j=1}^0$ (an empty sum) be equal to zero for convenience, consider the infinite expansion

$$\mathcal{L}_{\overline{F}}(s) = \sum_{k=0}^\infty s^{-1}\left(\exp\left\{\sum_{j=1}^k (-1)^j \frac{\kappa_j}{j} s^j\right\} - \exp\left\{\sum_{j=1}^{k+1} (-1)^j \frac{\kappa_j}{j} s^j\right\}\right), \tag{15}$$

and denote the kth term by the function $G_k(s)$. Each such function is actually of order b^k, and by carefully expanding them in an appropriate fashion, is the infinite sum of functions with known LT inverse. The initial term in the expansion is

$$G_0(s) = s^{-1}\left(1 - e^{-\kappa_1 s}\right) = \mathcal{L}_{1_{[0,\kappa_1]}}(s),$$

i.e., it is the LT of the indicator function on $[0, \kappa_1]$. This makes sense, because the right-tailed cdf should tend, as $b \to 0$, to an indicator function with boundary marked by its point mass, namely $\kappa_1 = \Lambda(0)$ (shown below). The higher order terms are more complicated, but contribute additional perturbations to this indicator function.

The key to the following theorem is the following class of polynomials: let ϕ_n be supported on $[0, \infty)$ such that

$$\psi_{n+1}(z, x) = \left(\frac{\partial^2}{\partial z \partial x}\left[\psi_n(z, x)\, e^{zx}\right]\right) e^{-zx}/n$$

and $\phi_n(x) = \psi_n(-1, x)$. Thus $\phi_1(x) = 1$, $\phi_2(x) = 1 - x$, $\phi_3(x) = 1 - 2x + x^2/2$, etc. These polynomials have the remarkable property that

$$\mathcal{L}_{\phi_{n+1}e^{-\cdot}}(s) = s^n(1 + s)^{-(n+1)}, \tag{16}$$

as shown in Gradshteyn and Ryzhik (1994). Now we can state the main expansion result, which applies more generally than to just the spectral density estimation problem.

Theorem 4. *Suppose that a Gaussian quadratic form $\langle Z, Z\rangle_T$ with covariance kernel K has cumulants given by (14). Then there exist coefficient sequences $\{\alpha_j^{(k)}\}$ for each $k \geq 1$ such that*

$$G_k(s) = \sum_{n=0}^\infty \frac{\alpha_{n+1}^{(k)}}{(n+1)!} s^n (1 + s)^{-(n+1)} = \sum_{n=0}^\infty \frac{\alpha_{n+1}^{(k)}}{(n+1)!} \mathcal{L}_{\phi_{n+1}e^{-\cdot}}(s)$$

and $G_0(s) = s^{-1}\left(1 - e^{-\kappa_1 s}\right)$, where $\sum_{k\geq 0} G_k$ is the Laplace Transform of the right-tailed cdf of $\langle Z, Z\rangle_T$. The right-tailed cdf has the expansion

$$\overline{F}(x) = 1_{[0,\kappa_1]}(x) + \sum_{n=0}^\infty \frac{\alpha_{n+1}^{(1)}}{(n+1)!} \phi_{n+1}(x) e^{-x} + \sum_{n=0}^\infty \frac{\alpha_{n+1}^{(2)}}{(n+1)!} \phi_{n+1}(x) e^{-x} + \cdots.$$

The coefficient sequences $\{\alpha_n^{(k)}\}$ are derived in the proof, and are fairly complicated expressions in terms of the cumulants. Next, we apply Theorem 4 to the case where $b \to 0$, noting that each subsequent term in the expansion is of higher order. As discussed in Sun et al. (2008) in the case of a regular taper and a short memory covariance kernel K, the cumulants satisfy $\kappa_j = O(b^{j-1})$; assuming this, we have the following corollary.

Corollary 1. *Suppose that a Gaussian quadratic form $\langle Z, Z\rangle_T$ with covariance kernel K has cumulants given by (14), and also suppose that $\kappa_j = O(b^{j-1})$ as $b \to 0$. Then*

$$\sum_{n=0}^\infty \frac{\alpha_{n+1}^{(k)}}{(n+1)!} \phi_{n+1}(x) e^{-x} = O(b^k)$$

as $b \to 0$, for each $k \geq 1$.

We note that the cumulants need not have the behavior $\kappa_j = O(b^{j-1})$ when long memory or negative memory is present, as demonstrated in McElroy and Politis (2012) for the $\theta = 0$ case. In that paper it was shown that the small b behavior of $S_0(b)$ has a distribution that either explodes to infinity (the case of long memory) or shrinks to zero (the case of negative memory).

Remark 2. As an example, consider the case that $\kappa_j = 0$ for $j > 2$, which corresponds to treating all higher order terms in b as zero. Then the LT of the pdf is just

$$\exp\{-\kappa_1 s + \kappa_2 s^2/2\},$$

which corresponds to a (positive) random variable with mean κ_1 and variance κ_2, and all higher order cumulants exactly zero. If the random variable were not enforced to be positive, it would correspond to the Gaussian distribution by its cumulant characterization. However, the actual limit is positive and non-Gaussian. Pretending – for the sake of making a comparison with the vanishing bandwidth fraction scenario – that this distribution is really Gaussian would yield the limit theorem

$$\frac{\widehat{f}(\theta)}{f(\theta)} \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(\kappa_1, \kappa_2).$$

The classic small-b results (Anderson, 1971) state that

$$\sqrt{N/M}\left(\frac{\widehat{f}(\lambda)}{f(\lambda)} - 1\right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, \int \Lambda^2(x)\, dx)$$

for $\lambda \in (0, \pi)$ and taper Λ (satisfying $\Lambda(0) = 1$) of bandwidth M, such that $M/N + 1/M \to 0$. Taking $M = bN$ in this result indicates that our results provide a higher order extension of the classical results, so long as $\kappa_1 = \Lambda(0)$ and $\kappa_2/b \sim \int \Lambda^2(x)\, dx$; this is shown in McElroy and Politis (2014).

Of course, these results are of a theoretical nature. The actual limiting quantiles vary dramatically with b and the long memory parameter β, so that the finite-sample distribution of the spectral density estimate is more accurately captured by utilizing both b and β, as opposed to using the vanishing bandwidth approximation via the normal distribution. This was demonstrated in simulation studies of HV for the $\beta = 0$ case; for $\beta \neq 0$, the situation is more egregious (using short memory quantiles when there is long/negative memory produces badly mis-sized statistics—see McElroy and Politis (2014)).

5. Numerical studies of size and bandwidth selection

This section now discusses some more practical aspects of spectral density estimation. We first discuss the simulated quantiles for the limit variable in Theorem 2 using three tapers, and investigate coverage in finite sample simulations as a function of bandwidth fraction b. A discussion of optimal bandwidth selection, as well as a more thorough exposition of quantile simulation and coverage, can be found in McElroy and Politis (2014).

5.1. Computing the spectral distribution

As in McElroy and Politis (2012), we have simulated the distribution of $S_\theta(b)$ for some tapers, when $\theta \neq 0, \pi$ (the case of $\theta = 0, \pi$ produces a distribution for $S_\theta(b)$ identical to the HAC case, and its quantiles can be found in published literature such as Kiefer and Vogelsang (2005)), and here we summarize the tables of McElroy and Politis (2014). We focus on the Bartlett and Trapezoidal tapers.

First consider the limit distribution $S_\theta(b)$ of Theorem 2 in the case that $\theta \neq 0, \pi$. In this case, recall that mean centering is irrelevant, so that the limit is a quadratic functional of FBM rather than FBB; moreover, there is a doubling effect, where $S_\theta(b)$ is really the sum of two iid random variables. In the case that $\theta = 0$ or $\theta = \pi$, the limit $S_\theta(b)$ is given by just one of these random variables. Furthermore, when $\theta = 0$ and we construct our spectral estimates by mean-centering, then the limit distribution $S_0(b)$ involves FBB instead of FBM. Alternatively, if no mean centering is utilized in the estimates (and the true mean is zero) then the distribution of $S_0(b)$ involves FBM rather than FBB. The case of $S_0(b)$ with mean-centering, utilized as a studentization of the sample mean, was studied in McElroy and Politis (2012); tables of quantiles for $S_0(b)$ and $S_\pi(b)$ have not been published, to our knowledge.

Both lower and upper quantiles have been estimated for the limit distributions, for a variety of long memory parameters β. (Specifically, in McElroy and Politis (2014) we consider $\beta = -.8, -.6, -.4, -.2, 0, .2, .4, .6, .8$, and the quantiles for $\alpha = .01, .025, .05, .1, .9, .95, .975, .99$.) Because the case of a frequency between 0 and π involves the sum of two iid variables (Theorem 2), versus just one such variable in the frequency 0 or π case, the quantiles in the former case are a bit larger and have more positive mass. When using non-positive definite tapers, such as the Trapezoidal tapers, the limit distribution has some mass on the negative half-line, and there is more such mass in the frequency 0 and π cases. No such negative mass occurs with the Bartlett taper, because it is positive definite. Another feature is the small b behavior of the quantiles as a function of β, namely that the quantiles shrink towards zero as β increases, when b is small. However, for negative memory ($\beta < 0$) the quantiles tend to decrease as a function of b, whereas the opposite is true for positive memory ($\beta > 0$). In comparison with the quantiles for the self-normalized sample means in the HAC literature (McElroy and Politis, 2009, 2011), the small b behavior as a function of memory is inverted, because in the self-normalized case the variable $S_b(0)$ is in the denominator of the limit distribution.

One implication of the small b behavior of S_b is that, while for short memory the distribution becomes centered around unity, for $\beta < 0$ there is more probability mass on values greater than one, whereas for $\beta > 0$ there is a greater probability of values less than one. When constructing confidence intervals, it is therefore possible in the long memory case for both the lower and upper quantiles to be less than unity, so that the confidence interval does not encase the estimator—the point estimate will lie below both the lower and upper limits of the interval. In contrast, for the negative memory case, the point estimate can lie above both the lower and upper limits of the interval. There is nothing incorrect mathematically about this feature, though it may look unusual (see the figures below for our applications); the effect diminishes as b is increased.

5.2. Simulation study of finite-sample coverage

The large bandwidth asymptotic theory provides a superior approximation to the finite-sample distribution of spectral estimators, as discussed in HV and Sun et al. (2008). Hence, this should provide superior coverage for confidence intervals and confidence bands; the work of HV illustrates this superior coverage, as compared to the classical normal approximation (utilizing small b methods). We seek here to extend those numerical results to an investigation of long memory, and also to spectral bands. Therefore, we first consider a seasonal long memory process CM($\beta, \pi/6$), adopting the pattern of study discussed in HV. Second, we consider an AR(2) process that generates a spectral peak, and compute the spectral distribution estimators, generating the corresponding confidence band. We are interested in determining the proportion of simulations for which the estimated spectral bands contain the true spectral distribution.

The long memory study begins by simulating 5000 Gaussian time series of length $N = 50, 100, 200$ from a process with spectral density

$$f(\lambda) = ([2 - 2\cos(\lambda - \theta)]\,[2 - 2\cos(\lambda + \theta)])^{-\beta},$$

which satisfies (3). Here we take $\theta = \pi/6$, which is a frequency of interest in monthly economic time series exhibiting seasonality (see Holan and McElroy (2012) for discussion of seasonal long memory modeling, and McElroy and Holan (2012) for computational aspects). As mentioned in Section 3, inference is conducted for the moving parameter $f_N(\theta)$, and we consider various values of the long memory parameter: $\beta = -.8, -.6, -.4, -.2, 0, .2, .4, .6, .8$. The case of $\beta = 0$ corresponds to white noise, and $f_N(\theta) = f(\theta)$ exactly – cf. (9) – in this case. Other values of $f_N(\pi/6)$ are produced in Table 1.

For each of 5000 simulations, we compute the spectral estimate $\widehat{f}(\theta)$ at the frequency $\theta = \pi/6$ of interest, construct the interval using (10), and record the proportion of simulations for which $f_N(\pi/6)$ is contained therein. We construct the interval using the true β (which would be unknown in practice) to determine quantiles. We consider two tapers (Bartlett and Trapezoidal .5) and four choices of $b = .04, .10, .20, .50$. This study differs somewhat from the approach in HV, which compares the cumulative distribution function (determined by simulation) of $N\widehat{f}(\theta)/V_N(\theta)$ against the cumulative distribution function of $S_b(\theta)$; here we highlight the incidence of under- or over-performance in finite sample. Of course, HV focuses on $\beta = 0$ as well. Tables 1 and 2 provide empirical coverage based on $\alpha = .05, .10$ confidence intervals, where the target quantity $f_N(\pi/6)$ is displayed in the tables as well.

The first thing to observe is the pattern of $f_N(\pi/6)$ as a function of N and β—it decreases with N if $\beta < 0$, is fixed at .5 when $\beta = 0$, and increases rapidly to infinity when $\beta > 0$. The actual coverage results display improved accuracy (in general) for increasing sample size, and somewhat inferior coverage for smaller values of b. Negative values of β lead to under-coverage (this can be quite poor for $\beta = -.8$ and small values of b), while positive values of β tend not to have worse performances, as compared to $\beta = 0$. In comparing tapers, it is quite noticeable that the Bartlett is inferior to the Trapezoidal taper; the latter is known to have superior bias properties in the short memory case, and this may be carrying over to the long memory case as well. Trapezoidal tapers tend to produce spectral estimates with slightly lower values than those constructed via a Bartlett taper, and the corresponding asymptotic distributions are shifted to the left (having positive mass at negative values). When a trapezoidal estimate produces a negative spectral estimate, the confidence interval construction given in Section 3 fails (because dividing by a negative number invalidates the inequality relations), and instead we replace such estimates with zero. (Another possibility is to replace negative values with half of the Bartlett estimate.) Essentially, our spectral estimate is computed using the maximum with zero, and the limit distribution should be modified accordingly. In cases where a lower quantile is negative, we replaced the lower boundary of the interval by zero (a more rigorous approach is to simulate the distribution $\max\{S_b(\theta), 0\}$); even using such an approximate technique, we obtained quite favorable results for the Trapezoidal taper, across all values of β.

Table 1
Empirical size for simulations of a cyclical long memory process of parameter $\beta = -.8, -.6, -.4, -.2, 0, .2, .4, .6, .8$, with spectral density estimates (at frequency $\pi/6$) computed using a Bartlett taper of bandwidth fraction $b = .04, .10, .20, .50$. Confidence intervals were constructed for $\alpha = .05, .10$, and empirical coverage is given in each cell for both nominal levels.

Empirical coverage for Bartlett taper									
β	$-.8$	$-.6$	$-.4$	$-.2$	0	.2	.4	.6	.8
$N = 50, f_{50}(\pi/6)$	.055	.084	.141	.256	.5	1.05	2.38	6.17	19.57
$b = .04$	.021, .017	.234, .187	.536, .445	.761, .670	.831, .744	.793, .706	.751, .661	.764, .658	.795, .706
$b = .10$	.233, .191	.614, .532	.809, .729	.885, .824	.915, .849	.916, .853	.925, .867	.932, .876	.939, .892
$b = .20$	.650, .569	.861, .788	.915, .858	.930, .877	.940, .890	.947, .896	.949, .901	.953, .904	.955, .908
$b = .50$	.786, .712	.904, .840	.930, .871	.943, .890	.946, .891	.948, .896	.948, .900	.947, .896	.947, .901
$N = 100, f_{100}(\pi/6)$	.033	.056	.108	.224	.5	1.20	3.14	9.30	33.36
$b = .04$	.022, .015	.333, .263	.687, .600	.852, .775	.884, .815	.880, .805	.873, .797	.897, .826	.917, .856
$b = .10$	.304, .251	.744, .663	.884, .805	.925, .866	.936, .882	.940, .885	.938, .884	.945, .890	.950, .905
$b = .20$	.732, .654	.906, .840	.930, .877	.942, .889	.952, .899	.943, .899	.954, .907	.957, .907	.959, .917
$b = .50$	.822, .753	.913, .854	.935, .883	.942, .896	.951, .901	.941, .886	.951, .901	.954, .901	.949, .904
$N = 200, f_{200}(\pi/6)$	.020	.038	.082	.195	.5	1.38	4.14	14.02	56.75
$b = .04$	.037, .025	.494, .396	.823, .738	.918, .855	.942, .882	.932, .877	.932, .872	.941, .887	.949, .903
$b = .10$	.366, .293	.812, .726	.920, .856	.942, .889	.949, .899	.945, .889	.942, .886	.944, .895	.951, .909
$b = .20$	.790, .725	.921, .863	.941, .888	.950, .895	.954, .903	.952, .900	.948, .897	.952, .906	.960, .920
$b = .50$	.848, .785	.930, .875	.947, .897	.949, .896	.948, .898	.948, .900	.944, .897	.949, .901	.952, .909

Table 2
Empirical size for simulations of a cyclical long memory process of parameter $\beta = -.8, -.6, -.4, -.2, 0, .2, .4, .6, .8$, with spectral density estimates (at frequency $\pi/6$) computed using a Trapezoidal .50 taper of bandwidth fraction $b = .04, .10, .20, .50$. Confidence intervals were constructed for $\alpha = .05, .10$, and empirical coverage is given in each cell for both nominal levels.

Empirical coverage for trapezoidal .50 taper									
β	$-.8$	$-.6$	$-.4$	$-.2$	0	.2	.4	.6	.8
$N = 50, f_{50}(\pi/6)$	.055	.084	.141	.256	.5	1.05	2.38	6.17	19.57
$b = .04$	.703, .652	.694, .631	.751, .697	.819, .750	.832, .758	.811, .731	.769, .676	.785, .687	.824, .740
$b = .10$	.870, .803	.933, .895	.952, .911	.952, .912	.950, .902	.957, .909	.955, .904	.958, .910	.965, .925
$b = .20$	.894, .797	.937, .867	.952, .895	.955, .904	.942, .892	.949, .898	.946, .892	.948, .893	.957, .906
$b = .50$	.904, .858	.925, .886	.942, .895	.941, .897	.950, .900	.947, .893	.945, .898	.941, .887	.951, .898
$N = 100, f_{100}(\pi/6)$	.033	.056	.108	.224	.5	1.20	3.13	9.30	33.36
$b = .04$	.918, .864	.944, .904	.938, .900	.938, .886	.927, .869	.928, .867	.933, .867	.942, .889	.954, .912
$b = .10$	.893, .834	.932, .903	.939, .899	.939, .890	.938, .884	.947, .892	.941, .887	.945, .887	.949, .901
$b = .20$	.917, .842	.956, .895	.954, .895	.951, .897	.955, .904	.952, .907	.984, .908	.954, .905	.960, .915
$b = .50$	.913, .866	.932, .891	.941, .900	.944, .895	.952, .899	.949, .896	.952, .900	.949, .896	.946, .900
$N = 200, f_{200}(\pi/6)$	.020	.038	.082	.195	.5	1.38	4.14	14.02	56.75
$b = .04$	.914, .853	.950, .908	.958, .923	.953, .913	.951, .906	.948, .898	.949, .895	.949, .896	.956, .914
$b = .10$	.920, .869	.937, .908	.944, .908	.947, .901	.949, .888	.945, .896	.951, .897	.946, .896	.953, .908
$b = .20$	.919, .844	.952, .886	.959, .898	.952, .905	.951, .897	.952, .907	.957, .908	.952, .903	.956, .912
$b = .50$	.927, .885	.943, .902	.943, .900	.948, .906	.948, .898	.948, .901	.950, .901	.945, .893	.950, .897

Now these coverage results are idealized, because we presume to know the true β when utilizing limit quantiles. In practice, an estimate of β would be obtained, and then appropriate quantiles could be simulated. If we instead always utilize $\beta = 0$ quantiles, even when mis-specified, the coverage deteriorates significantly (we have not systematically investigated this) because the quantile functions are quite sensitive to β.

For the second simulation study, we wish to investigate the coverage for the spectral distribution band method described in Section 3. We consider a cyclical process $\{Y_t\}$ given by the AR(2) equation $(1 - 2\rho\cos(\theta)B + \rho^2B^2)Y_t = \epsilon_t$, for $\{\epsilon_t\}$ a white noise process of unit variance. We consider $\theta = \pi/6$, and values $\rho = .7, .8, .9$ to generate several different cycles (values of θ closer to zero make estimation more challenging in some ways). Then we take 1000 Gaussian draws from this process of sample size $N = 50, 100, 200$, and compute the spectral distribution estimate $\widehat{G}(\cdot)$, and form the confidence band about it utilizing the quantiles ℓ and u from (13), determined using both the true unknown f to compute the kernel K (as an unavailable baseline) as well as the estimated $\widehat{K}$ utilizing an estimated spectral density (with mesh size $M = 600$), as described in Section 3.3. The spectral distribution estimate is constructed with a particular choice of taper (Bartlett or Trapezoidal .5) and bandwidth fraction ($b = .04, .1, .2, .5$), and the spectral density estimate used to estimate the kernel K uses the same specification. Once the bands are determined, the true spectral distribution can be plotted, and coverage is determined by the condition that the true function lies entirely within the bands. For $\alpha = .05, .10$ we determine the empirical coverage. Results are summarized in Tables 3 and 4; in each cell, the first entry corresponds to using the estimated $\widehat{K}$, whereas the second entry utilizes perfect knowledge of the true f.

It is remarkable that results for spectral band coverage are much inferior to those of spectral density coverage, in general. Under-coverage seems to be the general malaise, and small values of b accentuate the affliction. However, some intuitive results can be gleaned from the tables. First, coverage improves with sample size (albeit, sometimes moving from under-coverage to over-coverage); second, coverage is better for lower values of ρ; third, the under-coverage problem is less egregious when omniscience about f is utilized. Regarding the second point, recall that higher values of ρ indicate a sharper spectral peak, causing the spectral distribution to depart from a diagonal line (the case of white noise) and more closely resemble a step function—correctly capturing the width of uncertainty is more challenging when serial correlation is

Table 3
Empirical coverage for spectral distribution bands based on the Bartlett taper with bandwidth fraction $b = .04, .10, .20, .50$, based on simulations of sample size $N = 50, 100, 200$ from a cyclical AR(2) process of frequency $\theta = \pi/6$ and persistency $\rho = .7, .8, .9$. Coverage for both nominal levels $\alpha = .10, .05$ is given in each cell, with the first entry based on an estimator of the spectrum, and the second entry based on perfect knowledge of the true spectrum.

Empirical coverage for Bartlett taper						
Bandwidth	Coverage, $\alpha = .10$			Coverage, $\alpha = .05$		
$N = 50$	$\rho = .7$	$\rho = .8$	$\rho = .9$	$\rho = .7$	$\rho = .8$	$\rho = .9$
$b = .04$	.503, .917	.388, .886	.345, .942	.583, .974	.486, .956	.433, .972
$b = .10$	.768, .944	.703, .937	.647, .962	.811, .969	.761, .961	.707, .979
$b = .20$	.833, .952	.796, .942	.750, .966	.862, .969	.838, .961	.796, .979
$b = .50$	.856, .939	.851, .941	.818, .968	.889, .960	.887, .961	.848, .980
$N = 100$	$\rho = .7$	$\rho = .8$	$\rho = .9$	$\rho = .7$	$\rho = .8$	$\rho = .9$
$b = .04$	.703, .893	.642, .891	.428, .854	.765, .960	.716, .958	.516, .937
$b = .10$	.849, .946	.836, .948	.673, .942	.887, .977	.878, .980	.747, .963
$b = .20$	.884, .950	.877, .952	.785, .954	.913, .980	.913, .983	.832, .965
$b = .50$	.904, .946	.901, .954	.840, .956	.939, .976	.931, .982	.880, .967
$N = 200$	$\rho = .7$	$\rho = .8$	$\rho = .9$	$\rho = .7$	$\rho = .8$	$\rho = .9$
$b = .04$	.794, .896	.747, .886	.565, .806	.850, .935	.825, .942	.670, .903
$b = .10$	.883, .924	.882, .931	.803, .911	.919, .957	.918, .962	.850, .953
$b = .20$	.903, .927	.908, .937	.857, .925	.933, .958	.938, .965	.902, .962
$b = .50$	.924, .923	.932, .934	.897, .927	.943, .957	.956, .966	.926, .962

Table 4
Empirical coverage for spectral distribution bands based on the Trapezoidal .50 taper with bandwidth fraction $b = .04, .10, .20, .50$, based on simulations of sample size $N = 50, 100, 200$ from a cyclical AR(2) process of frequency $\theta = \pi/6$ and persistency $\rho = .7, .8, .9$. Coverage for both nominal levels $\alpha = .10, .05$ is given in each cell, with the first entry based on an estimator of the spectrum, and the second entry based on perfect knowledge of the true spectrum.

Empirical coverage for trapezoidal .50 taper						
Bandwidth	Coverage, $\alpha = .10$			Coverage, $\alpha = .05$		
$N = 50$	$\rho = .7$	$\rho = .8$	$\rho = .9$	$\rho = .7$	$\rho = .8$	$\rho = .9$
$b = .04$	.773, .904	.686, .913	.584, .921	.831, .931	.770, .939	.674, .939
$b = .10$	.861, .948	.839, .947	.754, .947	.879, .971	.874, .958	.797, .965
$b = .20$	.866, .949	.859, .949	.802, .949	.892, .973	.889, .960	.824, .968
$b = .50$	.883, .943	.878, .946	.825, .950	.914, .972	.905, .961	.855, .972
$N = 100$	$\rho = .7$	$\rho = .8$	$\rho = .9$	$\rho = .7$	$\rho = .8$	$\rho = .9$
$b = .04$	.888, .930	.868, .938	.755, .933	.915, .973	.902, .971	.823, .962
$b = .10$	.903, .949	.898, .949	.859, .945	.922, .975	.925, .974	.886, .967
$b = .20$	.912, .946	.911, .948	.880, .945	.930, .976	.937, .976	.908, .970
$b = .50$	.924, .935	.927, .943	.896, .945	.938, .972	.948, .976	.920, .969
$N = 200$	$\rho = .7$	$\rho = .8$	$\rho = .9$	$\rho = .7$	$\rho = .8$	$\rho = .9$
$b = .04$	.912, .943	.880, .925	.798, .910	.943, .977	.923, .963	.866, .953
$b = .10$	.916, .943	.909, .928	.902, .936	.948, .975	.941, .964	.921, .963
$b = .20$	.923, .938	.918, .928	.908, .932	.952, .971	.946, .959	.923, .960
$b = .50$	.943, .931	.935, .915	.918, .925	.965, .969	.962, .958	.935, .959

present. The third point has ramifications for hypothesis testing, where we might hypothesize a specific formula for f (e.g., white noise or a cyclical AR(2)) and then test this hypothesis by seeing whether the spectral distribution estimate is completely contained in bands computed from that particular f. Finally, the impact of taper can be seen with reduced under-coverage of the flat-top taper, which is especially prevalent in the small b case. However, we remind the reader that flat-top tapers are not positive definite, so that spectral density estimates can have negative values, and therefore the spectral distribution estimates need not be monotonically increasing in frequency.

6. Empirical applications

Spectral analysis has a diverse range of statistical applications; here we highlight a few applications through two empirical analyses.

6.1. Identifying residual seasonality in retail series

Suppose one is analyzing a monthly or quarterly economic time series, such as total retail sales, and is interested in identifying periodicities by estimating spectral peaks. For retail series, seasonality is typically the most salient dynamic in the data, effectively masking trend and business cycle movements. Seasonal adjustment should remove seasonality while ideally leaving the other dynamics undisturbed; inadequate seasonal adjustment is indicated by the presence of spectral peaks at seasonal frequencies. Here we utilize the spectral distribution function to affirm the adequacy of seasonal adjustment.

We apply the spectral distribution methodology to the monthly series of total retail sales for the major industry classifications 441 (Motor Vehicles and Parts Dealers), available from the US Census Bureau.[2] We consider a variety of tapers and bandwidth fractions for the seasonally adjusted data, covering the years 1992 through 2012. We will generate a graph of the spectral distribution estimate with sufficient resolution to examine business cycle effects, as well as seasonality, and also provide measures of uncertainty for the entire function. Because the business cycle has a period of two to ten years in general, the minimum number of frequencies needed

[2] Monthly Retail Trade and Food Services survey.

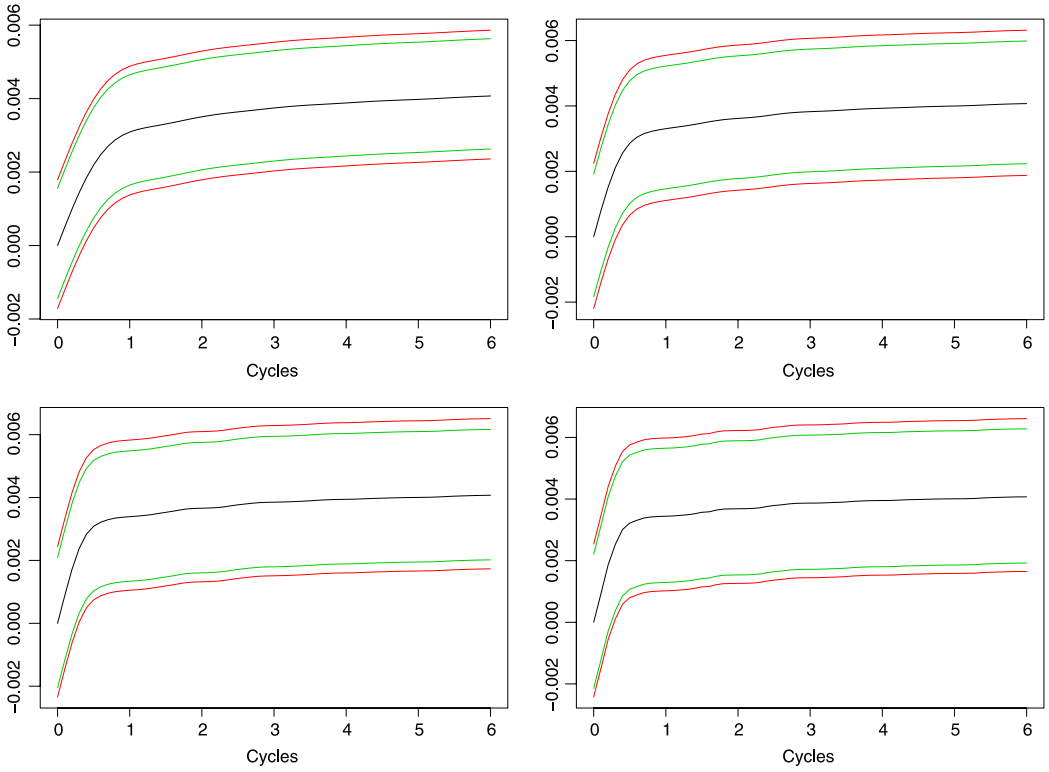

Fig. 1. Spectral distribution estimate using estimated covariance kernel for Retail 441 series, utilizing a Bartlett taper and bandwidth fractions $b = .04, .1, .2, .5$ in upper left, upper right, lower left, and lower right panels respectively. The confidence bands at .95 and .90 nominal coverage are displayed as red (outermost) and green (middle) curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is 60 (a ten year cycle for monthly data corresponds to frequency $\pi/60$). Thus we will take $\omega_j = \pi j/60$ for $0 \leq j \leq 60$; note that ω_j for $1 \leq j \leq 5$ are the business cycle frequencies. Also $\omega_0 = 0$ corresponds to the trend frequency, and ω_{10k} for $k = 1, 2, 3, 4, 5, 6$ corresponds to seasonal frequencies.

Supposing that this retail series has been processed by the program X-12-ARIMA, we then compute the (trend-differenced) adjustment's spectral distribution function, calculating the spectral bands to quantify uncertainty. Inadequate seasonal adjustment can result in distorted business cycle and/or trend estimates, and thus has severe implications for econometric analysis; see discussion in Bell and Hillmer (1984). Any sudden jumps around the seasonal frequencies indicate residual seasonality, while a straight diagonal line corresponds to perfect white noise. We compute the spectral distribution estimate and its confidence bands (with estimated kernel) utilizing the methodology of Section 5, considering both the Bartlett taper and the Trapezoidal .5 tapers, each with bandwidths $b = .04, .1, .2, .5$. The bands are computed at the range of frequencies discussed above, with green (middle) curves corresponding to the .90 coverage and red (outermost) curves for the .95 coverage.

Fig. 1 shows the result for the Bartlett taper, while Fig. 2 shows results for the Trapezoidal .50 taper. The steady growth in the spectral plots between cycles 0 and 1 (i.e., for frequencies up to $\pi/6$) indicates near constant spectral mass, and behavior similar to white noise; there is no sharp increase in the vicinity of any of the key seasonal cycles. One overall conclusion, from each of the plots, is that no significant seasonality remains. The impact of bandwidth fraction is much less apparent than in the spectral density estimates, which we expect from our asymptotic theory. One interesting feature can be discerned when comparing tapers; the trapezoidal tapers produce, in some cases, spectral distribution estimates that decrease at some frequencies, violating the fact that spectral distribution functions are monotonically increasing. This occurs because the flat-top tapers are not positive definite; in contrast the Bartlett taper, being positive definite, does not have this problem—though we can expect the width of the spectral bands about the estimator to be too small, especially for small b, as discussed in Section 5.

Since the spectral distribution displayed in Fig. 1 corresponds to the series that will subsequently be used for econometric analysis,[3] what can we deduce about Retail 441 series' dynamics? The most salient feature is a dramatic tapering in the slope somewhere around frequency $\pi/12$, or mid-way between cycles 0 and 1. This indicates a large degree of power at the trend and business cycle frequencies, which tapers off around $\pi/12$ with much less variation at higher frequencies. This is consistent with the generic spectral portrait of an economic series. Identifying the ordinate of this "kink" in the spectral distribution (which is more pronounced when b is larger) is delicate, depending on the taper and bandwidth, but can be used as the basis for designing band-pass filters for business cycle estimation.

6.2. Long memory spectral analysis of housing starts

Here we consider regional housing starts, for the South region, measured at a monthly frequency from 1964 through 2012, available from the US Census Bureau. We analyze the data here with a nonparametric approach, attempting to plot the spectral estimates for a variety of bandwidths, taking any seasonal long memory into account when quantifying uncertainty. We consider the same grid of frequencies as in the retail series, but are principally interested in the seasonal frequencies.

The South starts has been cleaned of outliers and level shifts, and we utilize a log transformation to stabilize variability.

[3] Typically, all economic analyses are based upon official published seasonal adjustments.

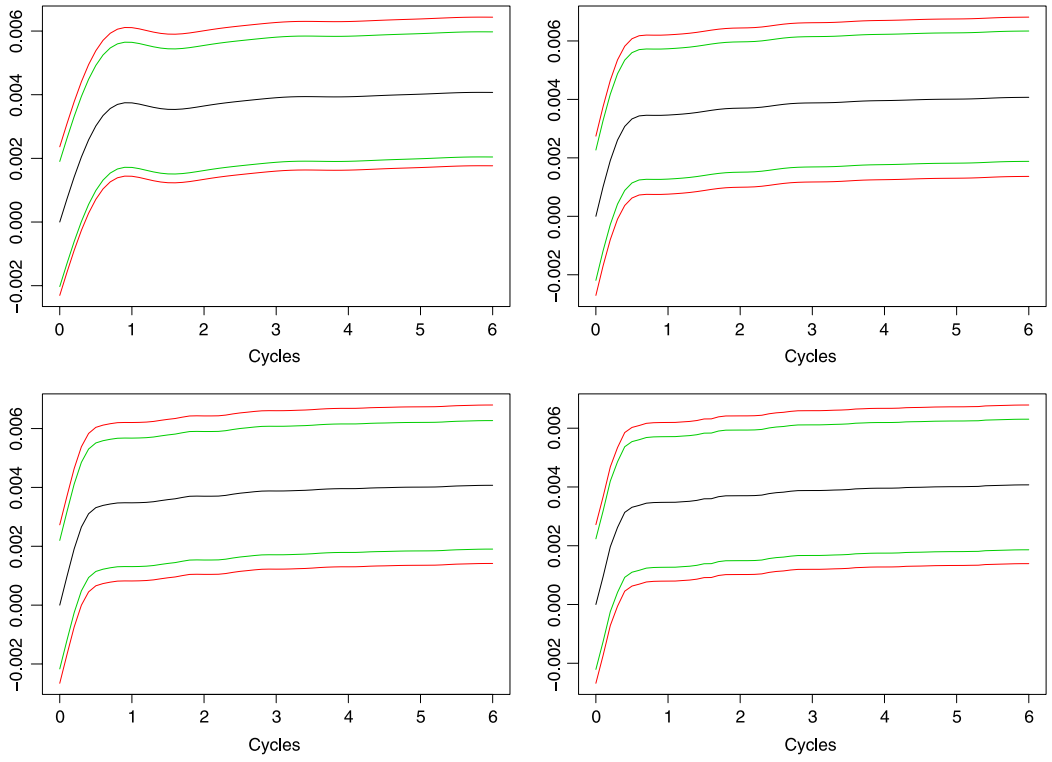

Fig. 2. Spectral distribution estimate using estimated covariance kernel for Retail 441 series, utilizing a Trapezoidal .50 taper and bandwidth fractions $b = .04, .1, .2, .5$ in upper left, upper right, lower left, and lower right panels respectively. The confidence bands at .95 and .90 nominal coverage are displayed as red (outermost) and green (middle) curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Analysis of sample autocorrelation plots for the first differences (to eliminate trend growth) reveals the presence of highly persistent correlation at seasonal lags (multiples of twelve), which indicates either nonstationarity or seasonal long memory. A common approach with such series is to utilize seasonal differencing – under the assumption of seasonal unit roots being present – but here we proceed with a hypothesis of stationarity, instead proceeding to estimate the seasonal long memory. This seems to be a plausible investigation, given the long sample size.

In order to obtain the right quantiles in each case, it is necessary to know the cyclical memory. It is reasonable to suppose, based on the form of nonstationarities in such series and the discussion above, that cyclical memory may be present at frequencies ω_{10k} for $0 \leq k \leq 6$, and at no others. To estimate the cyclical memory β_θ for these seven frequencies, one can adopt the crude estimation method described in McElroy and Politis (2011), adapted to nonzero frequencies:

$$\log\left(\widehat{f}(\theta)\right) = \beta \log(n) + \epsilon_n.$$

This regression equation is to be viewed as depending on sample size n, taking subsamples of length n for $100 \leq n \leq N$, with $N = 587$. The error ϵ_n is equal to the logarithm of the spectral estimate divided by n^β, and hence is approximately distributed as $\log S_b(\theta)$ when n is large. These regression errors are highly cross-correlated across various values of n, but nevertheless we will utilize ordinary least squares to get a rough estimate of β; see McElroy and Politis (2007) for a similar methodology.

We only need a rough estimate of β, because we only have quantiles for values of β belonging to the grid $\{-.8, -.6, -.4, -.2, 0, .2, .4, .6, .8\}$ anyways; we adopt the quantiles for a value of β closest to that derived from the regression. In this way we can obtain the quantiles for each spectral estimate at each of the six spectral peaks, using $\beta = 0$ at the non-seasonal frequencies. The estimates for the six spectral peaks are 0.42, 0.29, −0.02, 0.05, 0.09, and −0.01 respectively. Therefore we shall use $\beta = .4$ quantiles for the first peak, $\beta = .2$ quantiles for the second peak, and $\beta = 0$ quantiles for all other frequencies. We produce spectral density estimates only for the Bartlett taper (in log scale), with the bandwidth fractions $b = .04, .1, .2, .5$. (More extensive results can be found in McElroy and Politis (2014).) We focus on $\alpha = .05$, the results for $\alpha = .10$ looking quite similar. Recall that the quantiles utilized at frequency 0 and π are different, and induce a slightly wider interval; results are displayed in log scale for the Bartlett taper in Fig. 3.

The impact of bandwidth fraction is quite evident in the plots; smaller values of b enforce more smoothing. As was noted in Section 5, when long memory is present the confidence interval can lie completely above the point estimate, and this is evident in the figures with $b = .04$. Apart from the two long memory seasonal peaks, the other frequencies do not have this property, as they have short memory dynamics. We also highlight that at frequencies 0 and π the confidence intervals are slightly wider to reflect the heightened uncertainty. Note that the logarithmic transform, which was used for easier viewing, is not possible for the trapezoidal tapers, because the spectral density estimates take on negative values (not displayed). The impact of the trapezoidal taper, in contrast to the Bartlett, is to shift the estimate downwards—this improves bias and coverage, but at the cost of losing positivity. Otherwise, there is little to discriminate between the tapers, given the same choice of bandwidth.

The main features evident from this analysis are that two spectral peaks are present at the first and second seasonal frequencies. The evidence of the existence of spectral peaks at the third and fifth seasonal frequencies is less apparent, whereas there seems to be no peak at all at the fourth seasonal frequency. This summary is corroborated by the rough estimates of the cyclical memory parameters. As a result, we deduce that seasonal dynamics in South starts are driven by phenomena recurring once or twice a year. It is known that construction activity tapers considerably in the winter months, and weather predominantly drives the seasonality. Given

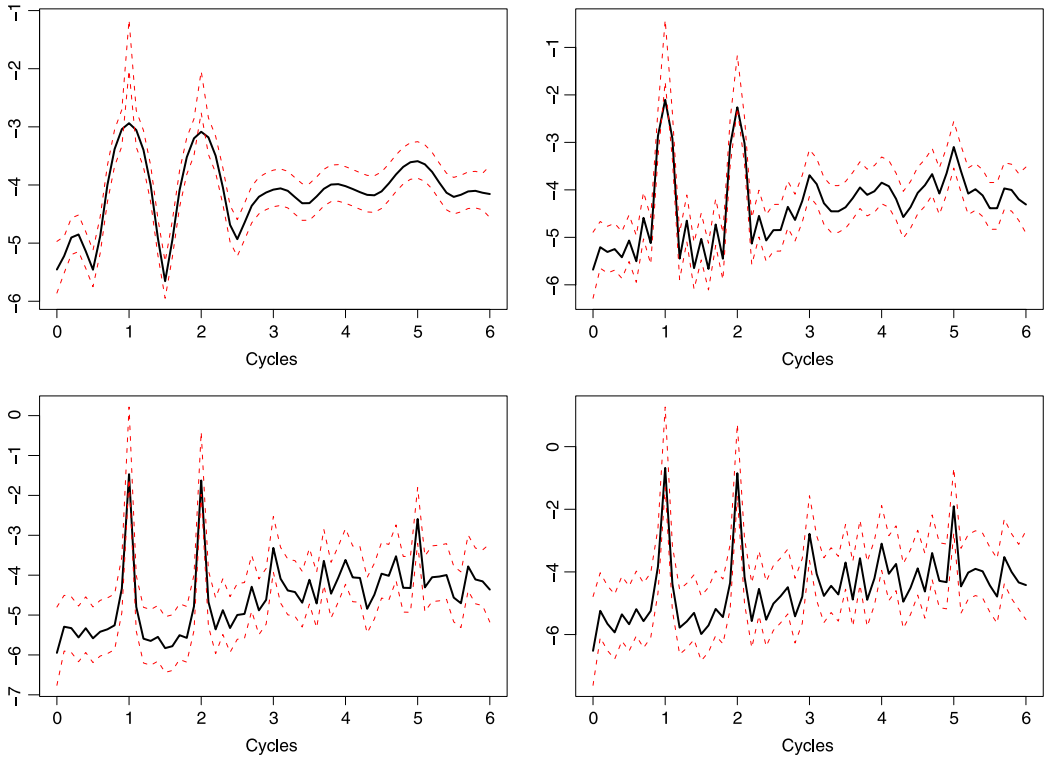

Fig. 3. Spectral density estimate in log scale for South series, utilizing a Bartlett taper and bandwidth fractions $b = .04, .1, .2, .5$ in upper left, upper right, lower left, and lower right panels respectively. The confidence intervals (red dashed lines) are for .95 nominal coverage.

this spectral portrait, signal extraction filters for the first and second stochastic seasonal could be derived or designed, and the once-a-year and twice-a-year persistent components could be estimated and separated. If the series is to be seasonally adjusted, a filter that suppresses only the first and second seasonal frequencies may be adequate.

7. Conclusion

This paper provides a new study of taper-based spectral estimation from the perspective of fixed bandwidth ratio asymptotics. Classical spectral estimation theory assumes that the bandwidth is negligible with respect to sample size, asymptotically, while the so-called "fixed-b asymptotics" allows for a constant ratio of bandwidth to sample size. Previous work on fixed-b asymptotics for spectral density estimation (HV) has focused on short memory dynamics and a single frequency, but we make extensions in several directions: (i) we study joint convergence over a finite collection of fixed frequencies; (ii) we allow for cyclical long memory at any of these frequencies; (iii) we provide results for flat-top tapers and tapers with kinks, extending the cases studied by HV (Bartlett and smooth tapers); (iv) we provide a discussion of higher-order accuracy in the short memory case, by an expansion of the cumulative distribution function of the spectral density estimate's limit; (v) we study spectral distribution estimation in the context of fixed-b asymptotics, and develop the application of simultaneous confidence bands; (vi) we tabulate the spectral density estimate's limit quantiles, as a function of taper, memory parameter, and bandwidth fraction; (vii) we empirically examine coverage of the spectral density and spectral distribution estimates.

The relevance of spectral estimation for econometric analysis is the following: the spectrum captures the second order structure of a stationary series in a compact, graphical manner. One may wish to stare at this graph to gain insight, as peaks in the spectral density (or steep increases in the spectral distribution) correspond to quasi-periodic dynamics in the data; this in turn indicates to the practitioner what types of filters can be utilized or designed to extract features of interest. Understanding these graphs can help one to avoid simple mistakes of generating spurious signal extraction estimates. For example, if one is interested in extracting a business cycle component and utilizes some sort of band-pass filter, a spurious cycle can be extracted from the data when no spectral peak actually exists—e.g., applying a band-pass filter to white noise will generate a business cycle from a purely random sequence. However, utilizing a band-pass filter when there is actually a spectral peak in that particular band will produce a reasonable estimate of actual cyclical dynamics.

As another example, we can design seasonal adjustment filters that have zeros only at the seasonal frequencies where actual peaks have occurred—recall the South starts application. By knowing where the peaks in the spectrum actually occur, we can avoid inducing unneeded zeros in the spectrum of the seasonal adjustment, which in turn can generate negative autocorrelation at seasonal lags in the filter output (see Bell and Hillmer (1984) for discussion). The broader point is that informed filter design flows naturally from inspection of the graph of a spectral density estimate. Many features of econometric interest – such as trends and business cycles and turning point indicators – can be captured through correct filter design.

For those econometricians disinclined to study spectral graphs, but are chiefly interested in the information conveyed by the parameters of fitted models, one may view the spectral density estimate as a proxy for a model-based spectrum, and essentially fit the model via matching the two. This is the philosophy behind the Whittle likelihood (see Taniguchi and Kakizawa (2000)), which aims to fit a time series model by choosing parameter values such that the integral (over all frequencies) of the ratio of spectral estimate (typically the periodogram, but taper-based estimates can also be used) to the model's spectral density is minimized (one also requires a penalty term involving the log innovation variance).

Yet another application would be the construction of the spectral distribution estimate from a fitted model's residuals, and a formal test (via the confidence band methodology described above) of the white noise hypothesis (which corresponds, leaving aside the impact of uncertainty in estimated parameters, to a correct model specification). Still other uses for the spectrum exist, which the interested reader can find in the abundant spectral literature.

Although this paper attempts to study several questions, many more are raised in the process. What is the statistical behavior, from a fixed-b perspective, when frequencies are becoming asymptotically closer to one another? What is a sensible criterion for locally-optimal bandwidth selection that takes into account the smoothness across multiple frequencies? (Thus, optimality should be discussed in different terms from the HAC literature, which only has a single frequency to consider.) Can the results here be utilized to improve long-memory estimation, extending the log-spectrum regression ideas of Geweke and Porter-Hudak (1983)? Some of these queries we plan to study in future research.

Disclaimer This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the US Census Bureau. Dimitris Politis' research was partially supported by NSF grant DMS 13-08319 and DMS 12-23137.

References

Anderson, T., 1971. The Statistical Analysis of Time Series. Wiley, New York.

Bell, W., Hillmer, S., 1984. Issues involved with the seasonal adjustment of economic time series. J. Bus. Econom. Statist. 2, 291–320.

Blackman, R.B., Tukey, J.W., 1959. The Measurement of Power Spectra from the Point of View of Communications Engineering. Dover, New York.

Bohman, H., 1960. Approximate Fourier analysis of distribution functions. Ark. Mat. 4, 99–157.

Boutahar, M., 2008. Identification of persistent cycles in non-Gaussian long memory time series. J. Time Ser. Anal. 29, 653–672.

Dahlhaus, R., 1985. Asymptotic normality of spectral estimates. J. Multivariate Anal. 16, 312–431.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen, B.C., 1998. New capabilities and methods of the X-12-ARIMA seasonal adjustment program. J. Bus. Econom. Statist. 16, 127–177. with discussion.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. J. Time Ser. Anal. 4, 221–238.

Gradshteyn, I., Ryzhik, I., 1994. Table of Integrals, Series, and Products. Academic Press, New York.

Gray, H., Zhang, N., Woodward, W., 1989. On generalized fractional processes. J. Time Ser. Anal. 10, 233–257.

Grenander, U., Rosenblatt, M., 1953. Statistical spectral analysis arising from stationary stochastic processes. Ann. Math. Statist. 24, 537–558.

Hashimzade, N., Vogelsang, T., 2008. Fixed-b asymptotic approximation of the sampling behaviour of nonparametric spectral density estimators. J. Time Ser. Anal. 29, 142–162.

Holan, S., McElroy, T., 2012. Bayesian seasonal adjustment of long memory time series. In: Bell, W., Holan, S., McElroy., T. (Eds.), Economic Time Series: Modeling and Seasonality. Chapman and Hall, New York.

Karatzas, I., Shreve, S., 1991. Brownian Motion and Stochastic Calculus. Springer, New York.

Kiefer, N., Vogelsang, T., 2002. Heteroscedastic-autocorrelation robust standard errors using the Bartlett kernel without truncation. Econometrica 70, 2093–2095.

Kiefer, N., Vogelsang, T., 2005. A new asymptotic theory for heteroscedasticity-autocorrelation robust tests. Econometric Theory 21, 1130–1164.

Kiefer, N., Vogelsang, T., Bunzel, H., 2000. Simple robust testing of regression hypotheses. Econometrica 68, 695–714.

McElroy, T., Holan, S., 2012. On the computation of autocovariances for generalized Gegenbauer processes. Statist. Sinica 22, 1661–1687.

McElroy, T., Politis, D., 2007. Computer-intensive rate estimation, diverging statistics, and scanning. Ann. Statist. 35, 1827–1848.

McElroy, T., Politis, D., 2009. Fixed-b Asymptotics for the Studentized Mean from Time Series with Short, Long or Negative Memory. Department of Economics, UCSD, UC San Diego.

McElroy, T., Politis, D., 2011. Fixed-b Asymptotics for the Studentized Mean for Long and Negative Memory Time Series. Department of Economics, UCSD, UC San Diego.

McElroy, T., Politis, D., 2012. Fixed-b asymptotics for the studentized mean from time series with short, long or negative memory. Econometric Theory 28, 471–481.

McElroy, T., Politis, D., 2014. Spectral Density and Spectral Distribution Inference for Long Memory Time Series via Fixed-b Asymptotics. Department of Economics, UCSD, UC San Diego, http://escholarship.org/uc/item/6164c110.

Parzen, E., 1957. On consistent estimates of the spectrum of a stationary time series. Ann. Math. Statist. 28, 329–348.

Phillips, P., 1998. New tools for understanding spurious regressions. Econometrica 66, 1299–1325.

Phillips, P., Sun, Y., Jin, S., 2006. Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. Internat. Econom. Rev. 47, 837–894.

Politis, D., 2001. On nonparametric function estimation with infinite-order flat-top kernels. In: Charalambides, Ch., et al. (Eds.), Probability and Statistical Models with Aapplications. Chapman and Hall/CRC, Boca Raton, pp. 469–483.

Politis, D., Romano, J., 1995. Bias-corrected nonparametric spectral estimation. J. Time Ser. Anal. 16, 67–103.

Politis, D., Romano, J., You, L., 1993. Uniform confidence bands for the spectrum based on subsamples. In Tarter, M., Lock, M. (Eds.) "Computing Science and Statistics, Proceedings of the 25th Symposium on the Interface", San Diego, CA, April 14–17, 1993. *The Interface Foundation of North America*, pp. 346–351.

Priestley, M.B., 1981. Spectral Analysis and Time Series. Adademic Press, New York.

Samorodnitsky, G., Taqqu, M., 1994. Stable Non-Gaussian Random Processes. Chapman & Hall, New York City, New York.

Sun, Y., Phillips, P., Jin, S., 2008. Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. Econometrica 76, 175–194.

Taniguchi, M., Kakizawa, Y., 2000. Asymptotic Theory of Statistical Inference for Time Series. Springer-Verlag, New York City, New York.

Taqqu, M., 1975. Weak convergence to fractional Brownian motion and to the Rosenblatt process. Z. Wahrscheinlichkeitstheor. Verwandte Geb. 31, 287–302.

Tziritas, G., 1987. On the distribution of positive-definite Gaussian quadratic forms. IEEE Trans. Inform. Theory 33, 895–906.

Velasco, C., Robinson, P., 2001. Edgeworth expansions for spectral density estimates and studentized sample mean. Econometric Theory 17, 497–539.

Woodroofe, M., Van Ness, J., 1967. The maximum deviation of sample spectral densities. Ann. Statist. 38, 1558–1569.

Woodward, W., Cheng, Q., Gray, H., 1998. A k-factor GARMA long memory model. J. Time Ser. Anal. 19, 485–504.