RESEARCH REPORT SERIES
(Statistics #2017-01)

Detection of Seasonality in the Frequency Domain

Tucker McElroy
Anindya Roy

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: January 11, 2017

Detection of Seasonality in the Frequency Domain

Tucker McElroy[1] and Anindya Roy[1,2*]
[1]U.S. Census Bureau, [2]University of Maryland Baltimore County, Baltimore, MD, USA

Summary

Seasonal adjustment methods process and publish millions of time series across the world each month, and judgment of the adequacy relies heavily upon frequency domain diagnostics. In particular, peaks in the spectral density estimates of seasonally adjusted data are indicative of an inadequate adjustment. Spectral peaks are currently assessed in the X-12-ARIMA program via the visual significance approach, but this method lacks a rigorous statistical foundation. This paper provides such a foundation by providing measures of uncertainty for spectral peak measures, allowing for formal hypothesis testing. To apply the test we develop fixed-bandwidth fraction asymptotics for taper-based spectral density estimates.

Keywords: Fixed-b asymptotics; Seasonal adjustment; Spectral peaks; Visual significance;

1 Introduction

The detection and estimation of peaks in the spectral density of a stationary time series is a problem of long-standing interest, and is vital in the field of seasonal adjustment. Quarterly or monthly economic time series typically exhibit seasonality, most often described via a nonstationary stochastic process with unit root frequencies corresponding to the known seasonal frequencies (Bell and Hillmer, 1983). Adequate estimation and removal of seasonality should correspond to the absence of spectral peaks at these same seasonal frequencies in the adjusted series, and therefore the detection of spectral peaks is an important diagnostic for the adequacy of a seasonal adjustment (Findley, 2005). See the discussion in Chapter 3 of Hylleberg (1986). The problem is of widespread importance, because millions of time series are seasonally adjusted each month at statistical agencies around the world, many of whom utilize the software program X-12-ARIMA of the U.S. Census Bureau.

*Corresponding author: Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, anindya@umbc.edu

One approach to the detection of seasonality is to postulate as a null hypothesis the existence of a sinusoid – corresponding to a deterministic seasonal component – at the frequency of interest, and test whether spectral density estimates warrant such a hypothesis. The early literature on spectral peak testing (see Priestley (1981)) focused on this approach, and the stable-F test approach of Lytras et.al. (2007) does as well. Any stochastic seasonal component, conceived as a nonstationary process, can include such a stable sinusoidal component without loss of generality – this is analogous to the fact that a random walk with drift can be decomposed into a linear term (its mean) plus the purely stochastic mean zero portion. Tests for stable sinusoids focus on the deterministic part of seasonality, but are not designed to address the stochastic portion. However, it is important to do so: removal of the deterministic portion alone (say, via regression) does not entail the removal of the whole stochastic seasonal, and such an approach fails to accomplish the goal of seasonal adjustment – for most economic series, the seasonality is too evolutive to be adequately captured by fixed periodic functions.

Therefore, detection of seasonality should not be restricted to consideration of deterministic seasonality alone, but of the full stochastic nature of seasonality. Early work on assessing the effect of seasonal adjustment appeared in Nerlove (1964) and Grether and Nerlove (1970). Pierce (1976, 1979) looked at adequacy of seasonal adjustment by looking at the magnitude of the autocorrelations at seasonal lags of the adjusted series. This approach is generalized to the Q_s statistics, adopted by TRMAO-SEATS (Maravall, 2012) and discussed later in the data analysis section.

Whereas a deterministic sinusoid corresponds to a level shift in the spectral distribution function – and will appear in a spectral density estimate as a tall slender peak – stochastic seasonality instead corresponds to a broader peak in the spectral density estimate (e.g., computed using an Autoregressive Estimator); it will have a broader peak that nonetheless is approaching an infinite height as sample size increases. It becomes important to consider the width of a spectral peak in its assessment. Since the procedure of seasonal adjustment, viewed in the frequency domain, amounts to multiplication of a function with a peak by another function with a trough, whether or not the peak is transformed into a trough or not depends on the width of these functions (this point is discussed further in Section 2). For this and other reasons, Soukup and Findley (1999) considered a measure of the peak that involved the distance between ordinates of the log spectrum when examined on a grid of frequencies of mesh size $\pi/60$.

The actual measure proposed in Soukup and Findley (1999) – which has now become somewhat of a standard by virtue of its incorporation into the X-12-ARIMA software used at most international statistical agencies – is computed by comparing the spectral peak ordinate to both nearest neighbor ordinates, with respect to the chosen frequency grid; when both ordinate differences exceed a threshold (selected based upon empirical criteria), the spectral peak is declared to be "visually significant." So far no distribution theory has been proposed for this statistic, making it difficult to rigorously determine Type I and II error. As the null hypothesis in this situation says that a

peak actually does exist (where the definition of a peak depends upon the threshold), Type I error corresponds to declaring a seasonal adjustment as adequate when stochastic seasonality actually remains, whereas Type II error corresponds to falsely identifying spurious residual seasonality – this can happen because random perturbations in the spectral density can yield the appearance of a peak, even when the underlying time series has no such periodicity present. This paper proposes to put the concept of visual significance upon a rigorous statistical footing.

In order to develop a statistical theory for spectral peaks, one first requires a theory for spectral density estimation. Spectral density estimates generally fall into two classes: model-based (e.g., the Autoregressive spectral estimator, or other estimators derived from a fitted model) and nonparametric (e.g., based on smoothing the periodogram). We focus on the latter class, based upon tapering the sample autocovariances with a positive definite taper, such as the Bartlett or Daniell kernels. Asymptotic theory for such estimates goes back to Parzen (1957), and the literature adopts the perspective that the taper bandwidth is negligible relative to sample size. More recent literature, as in Hashimzade and Vogelsang (2008), adopts the perspective of the so-called fixed-b asymptotics, where the ratio of bandwidth length to sample size is assumed to be a fixed fraction $b \in (0, 1)$. Because the fixed-b asymptotic framework has several advantages – including a superior approximation of the sampling distribution, further discussed in McElroy and Politis (2014) – we pursue spectral peak detection with this perspective in mind.

Alternatively, there may be an advantage to assessing residual seasonality through the spectral distribution function. This function is well-defined for a broader class of time series, just as the cumulative distribution function is more broadly well-defined than the probability density function. Also, statistical estimation is somewhat simpler, with conventional central limit theory being applicable (Dahlhaus 1983, 1985). Because the spectral distribution is the anti-derivative of the spectral density, a peak in the latter corresponds to a steep rise in the former (which is monotonically increasing). In this article we only analyze peak measures that are based on the spectral density estimates.

One of the subtleties of this topic is the very definition of peak, which must be considered relative to neighboring peaks and troughs, relative to the overall scale, and in terms of the steepness of the peak. Given three values of a function, wherein the middle value exceeds its two close neighbors, we can describe two right triangles – adjoining one another – whose angles determine the steepness of the peak. Hence it can be deduced that the spacing of the two neighbors relative to the middle ordinate is important, and is related to how great the middle value must exceed the neighboring values to be judged a true peak. Such criterion has been proposed – via the visual significance concept – but we provide a more thorough mathematical basis for the notion of spectral peak.

This paper begins with some preliminaries in Section 2, where we describe a class of peaked Autoregressive spectra in order to provide a context for the discussion whether the visual significance estimand actually conveys the notion of "peakedness" i.e., whether the visual significance

measure is a reasonable proxy for the concept of spectral peak. We also demonstrate explicitly how spectral peaks, when filtered by a seasonal adjustment filter of varying signal-to-noise ratio, may or may not result in adequate seasonal adjustments. We further provide discussion of other possible notions of spectral peak – through the concept of so-called peak functionals. Finally, to understand and study the behavior of testing procedures based on peak functionals we introduce a new class of spectral densities that parameterize the process in terms of relevant features of the peak.

Section 3 describes our statistical methodology, whereby we consider the tapered spectral density estimators when the bandwidth length of the taper relative to the sample size is a constant fraction. Some of the results require extensions of previous literature, such as Hashimzade and Vogelsong (2008) and McElroy and Politis (2014); we allow the frequencies of interest to depend on sample size, and can be more general than Fourier frequencies. Then Section 4 treats the application of these estimators to peak detection through a hypothesis testing framework. Section 5 provides the size and power results of simulation studies, and Section 6 applies the methods to U.S. Census Bureau time series, including retail, construction, and manufacturing sectors. Section 7 concludes, and proofs are in the Appendix.

2 Spectral Peaks and Seasonal Adjustment

Here we discuss the concept of a spectral peak in the context of parametric classes of spectral densities. The primary objective is to gain insight for a hypothesis testing framework based on visual significance measure, particularly for formulation of a null hypothesis that accurately reflects the visual perception of absence of a peak.

2.1 The Seasonal Adjustment Motivation for Peak Detection

An example of a stationary time series $\{X_t\}$ with a single spectral peak is given by an AR(2) with $\phi_1 = -2\rho cos\theta$, $\phi_2 = -\rho^2 (0 < \rho < 1)$, so $\{X_t\}$ satisfies

$$(1 - 2\rho\cos(\theta)B + \rho^2 B^2)X_t = \epsilon_t, \tag{1}$$

with $\{\epsilon_t\}$ a white noise of variance σ^2. The autocorrelations of such series decay as a damped sine wave whose damping factor is ρ (see Box and Jenkins 1976, p. 60), whence its name persistency.

As $\rho \to 1$, the maximum value of $f(\lambda)$, given by $\arccos[(1+\rho^2)\cos(\theta/(2\rho))]$, tends to θ. Thus, θ effectively governs peak strength for values of ρ considered, $0.9 < \rho < 1$.

The spectral density for $\{X_t\}$ is

$$f(\lambda) = \frac{\sigma^2}{|1 - 2\rho\cos(\theta)e^{-i\lambda} + \rho^2 e^{-i2\lambda}|^2} = \frac{\sigma^2}{(1 - 2\rho\cos(\theta+\lambda) + \rho^2)\ (1 - 2\rho\cos(\theta-\lambda) + \rho^2)}, \tag{2}$$

which is maximized at $\lambda_0 = \arccos[(1+\rho^2)\cos(\theta)/(2\rho)]$ (when this quantity exists). Suppose that θ represents a seasonal frequency, e.g., $\theta = j\pi/6$ for $1 \leq j \leq 6$ are the seasonal frequencies for a monthly series. In Figures 1, 2, and 3 we display spectral densities in log scale (with $\sigma = 1$) for the seasonal frequencies of $\pi/6$, $2\pi/6$, and $3\pi/6$, respectively. The various panels display different values of ρ, moving from a less pronounced peak with $\rho = .9$, to a sharp peak where $\rho = .99$.

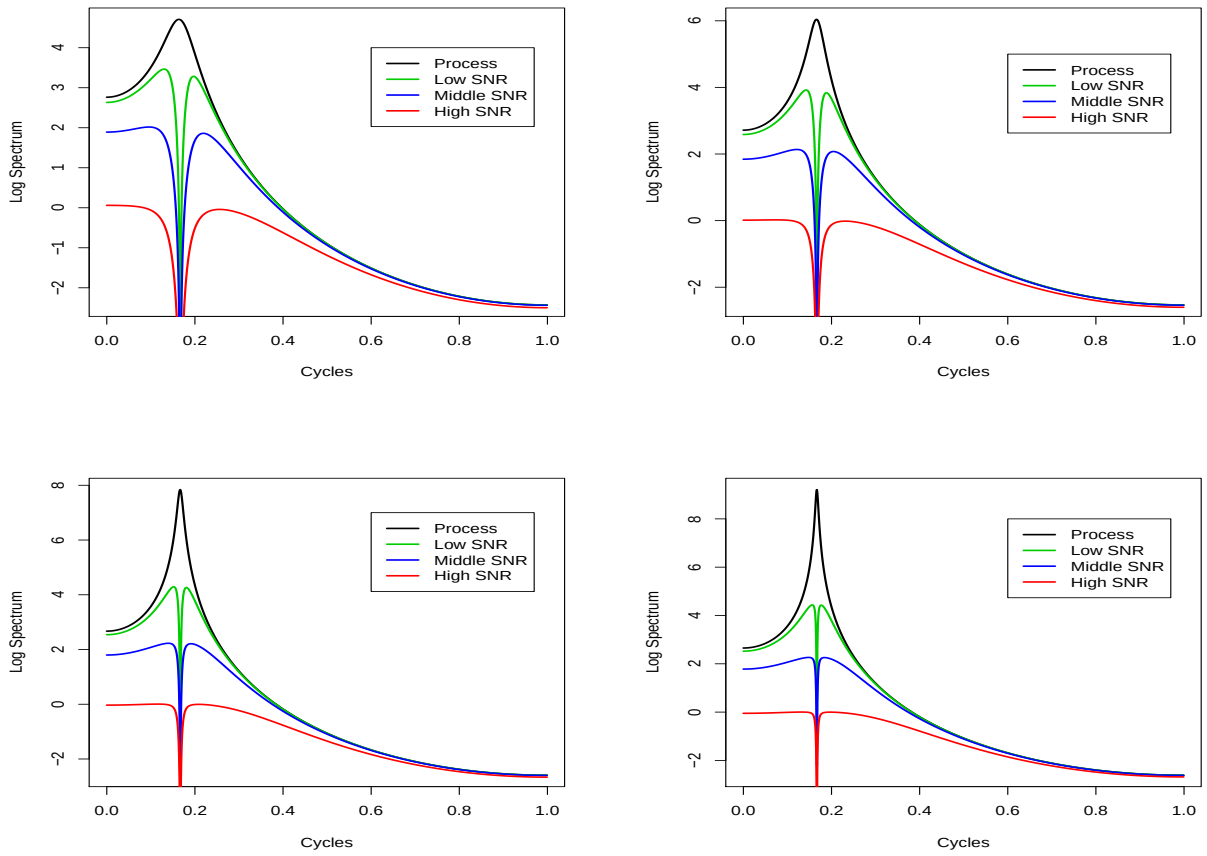


Figure 1: Log spectral densities for AR(2) process with $\omega = \pi/6$ (black) with seasonal adjusted log spectra in red (high signal-to-noise ratio), blue (medium signal-to-noise ratio), and green (low signal-to-noise ratio). Upper left panel corresponds to $\rho = .9$, upper right panel to $\rho = .95$, lower left panel to $\rho = .98$, and lower right panel to $\rho = .99$.

Seasonal adjustment typically proceeds by application of a filter $\Psi(B)$ to the input time series $\{X_t\}$, which results in an output series $\{Y_t\}$ such that $Y_t = \Psi(B)X_t$. It is well-known (Priestley, 1981) that if the input time series is stationary and non-deterministic, the spectral density of the output series is given by $|\Psi(e^{-i\lambda})|^2$ (the filter's squared gain function) times the spectral density of $\{X_t\}$. For purposes of illustration, let us suppose that the input time series has spectrum (2),

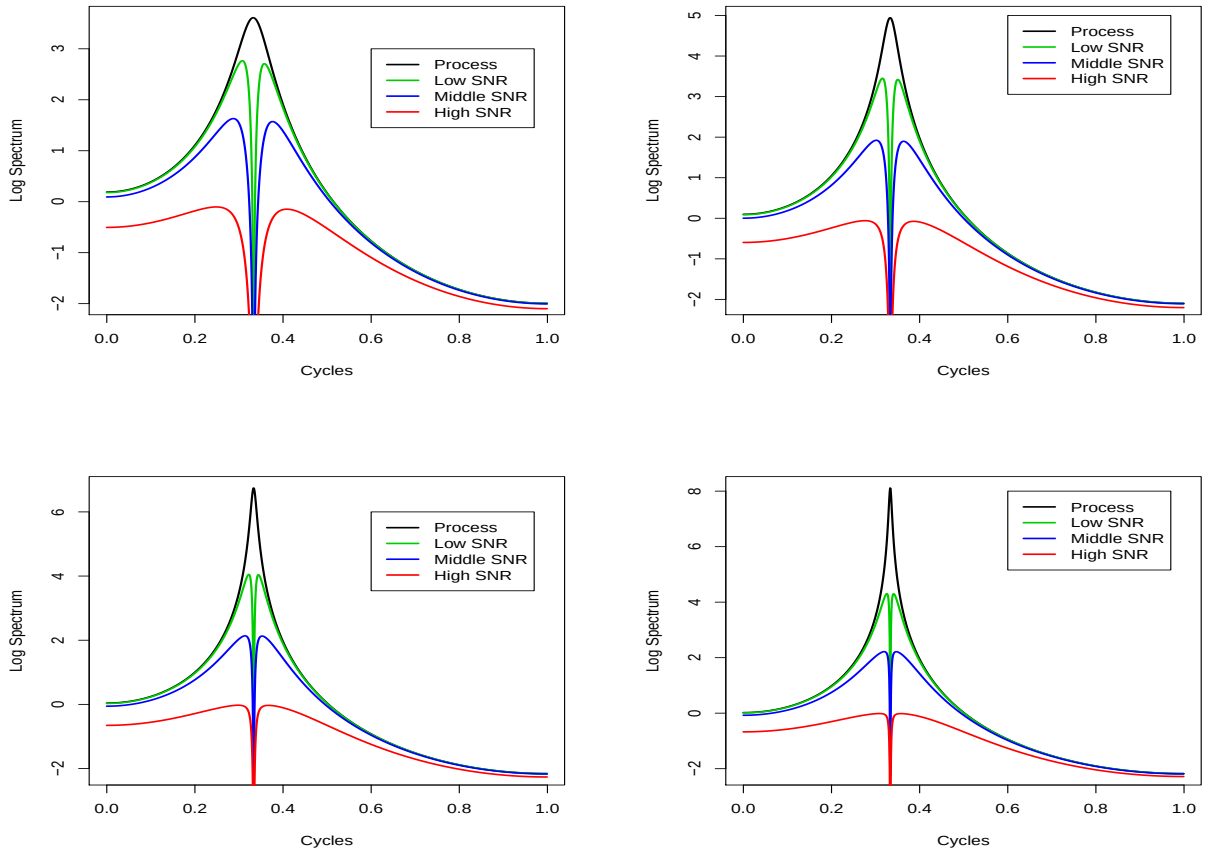


Figure 2: Log spectral densities for AR(2) process with $\omega = 2\pi/6$ (black) with seasonal adjusted log spectra in red (high signal-to-noise ratio), blue (medium signal-to-noise ratio), and green (low signal-to-noise ratio). Upper left panel corresponds to $\rho = .9$, upper right panel to $\rho = .95$, lower left panel to $\rho = .98$, and lower right panel to $\rho = .99$.

and that the seasonal adjustment filter has squared gain function given by

$$G(\lambda; q) = \frac{(1 - 2\cos(\theta + \lambda) + 1)\ (1 - 2\cos(\theta - \lambda) + 1)}{q^{-1} + (1 - 2\cos(\theta + \lambda) + 1)\ (1 - 2\cos(\theta - \lambda) + 1)},$$

where q is a signal-to-noise ratio (SNR) parameter. This filter is the Wiener-Kolmogorov seasonal adjustment filter for a process given by white noise plus a latent seasonal of spectrum (2), where $\rho = 1$. Such a filter produces a trough at frequencies near θ, and $G(\theta; q) = 0$. Higher values of q make this trough more slender, and are associated with less smoothing, while lower values of q make the trough broader. When the seasonality has high variability – relative to ambient noise – then the signal-to-noise ratio is high, and less smoothing is necessary, the signal (i.e., the seasonality) already being apparent; conversely, when seasonality has low variability more smoothing is needed to extract it.

Plotted in Figures 1, 2, and 3 are the product of the squared gain function with $f(\lambda)$, with

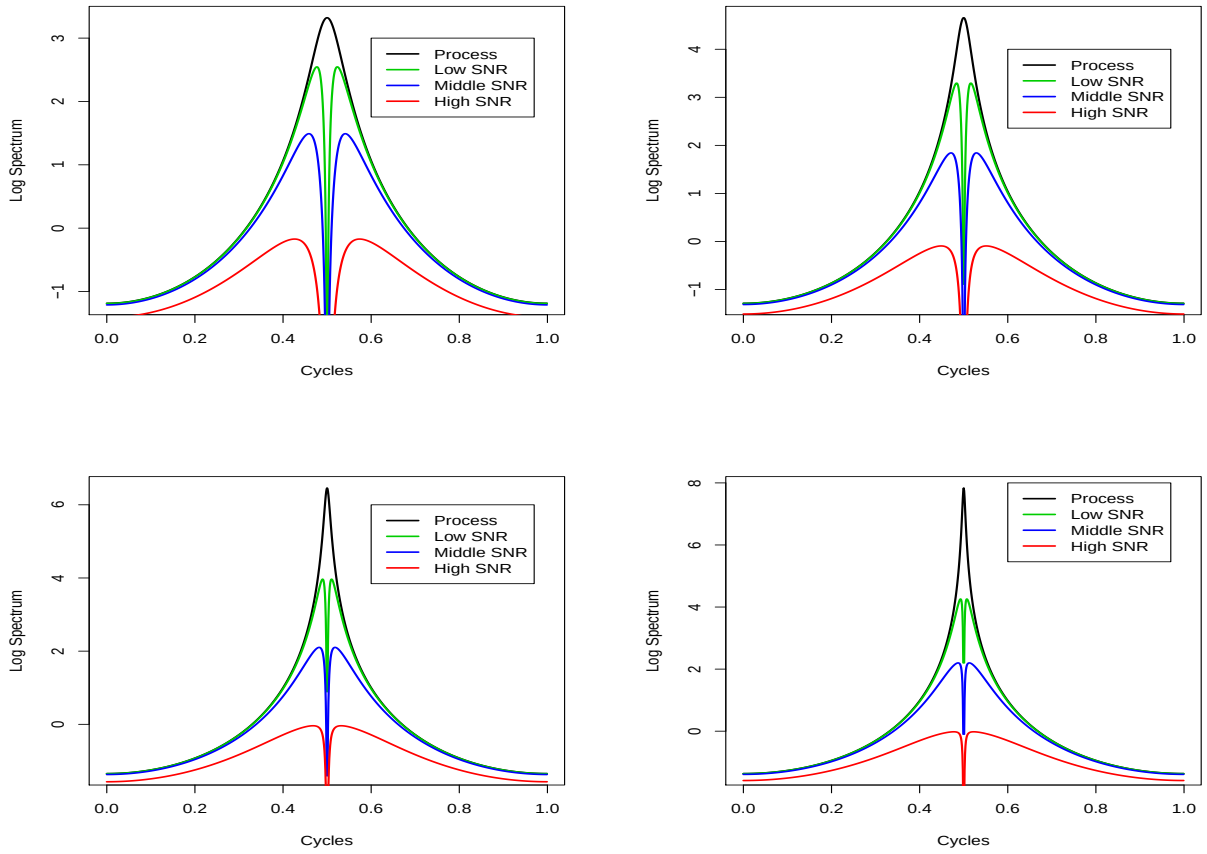


Figure 3: Log spectral densities for AR(2) process with $\omega = 3\pi/6$ (black) with seasonal adjusted log spectra in red (high signal-to-noise ratio), blue (medium signal-to-noise ratio), and green (low signal-to-noise ratio). Upper left panel corresponds to $\rho = .9$, upper right panel to $\rho = .95$, lower left panel to $\rho = .98$, and lower right panel to $\rho = .99$.

$q = 100, 10, 1$ in green, blue, and red, respectively. When q is high much of the overall peak shaped is retained, but with a slender trough in the center – producing what some authors refer to as an "inverted W shape." Decreasing q has the effect of leveling the peak, and eventually the trough pattern consumes, as it were, the peak. The effect also depends upon θ and ρ.

The point of these examples is the following. Suppose we seasonally adjust using a value of q that is too high for the data, resulting in an output spectral density with shape at the seasonal frequencies given by the green lines, corresponding to high SNR. If we tested for residual seasonality by only looking for fixed effects, we would likely wrongly conclude the adjustment to be adequate, because the slender trough indicates that all deterministic seasonality has been annihilated (the numerator of $G(\lambda; q)$ corresponds to the squared gain function of a second order differencing polynomial $1 - 2\cos(\theta)B + B^2$ which being a factor of $\Psi(B)$ ensures that $\Psi(B)$ annihilates all function s of period π/θ). A frequency-domain based test statistic, on the other hand, might be capable

of identifying the inadequacy. Such a test statistic should be an estimator of a parameter – a functional of the spectral density – that appropriately conveys the essence of being a peak.

2.2 Measuring Spectral Peaks

The idea of a peak in a graph is surprisingly subtle to capture through mathematical formulas. McElroy and Holan (2009) set forth a measure based upon measuring the second derivative of the spectral density, and other measures have been set forth. The approach of Soukup and Findley (1999) compares the log Autoregressive (AR) spectrum at a frequency of interest θ (in units of radians), e.g., a seasonal or trading day frequency, to two "nearest neighbors" on the left and right, some distance δ away. That is, we have three frequencies $\theta - \delta, \theta, \theta + \delta$, and comparisons are given by $\log f(\theta) - \log f(\theta - \delta)$ and $\log f(\theta) - \log f(\theta + \delta)$ for the left and right hand respectively. Note that no peak is present if one of these differences is non-positive. But even when both differences are positive, the actual shape of the peak may be so mild as to be indistinguishable from the overall shape of the spectral graph. Clearly this also depends on the spread between the three frequencies as well.

The Visual Significance approach of Soukup and Findley (1999) is to first fix a fairly uniform grid by dividing $[0, \pi]$ into 61 frequencies. We use the qualifier "fairly," since the trading day frequencies can also be considered, and do not fall exactly on the grid points $\pi j/60$, although the seasonal frequencies certainly do. Thus the distance to the nearest neighbors is $\delta = 1/60 \approx 1.7\%$ of the total width. These differences – computed in terms of the log spectrum $\log f$ – must both be above a pre-specified threshold τ_f, which is taken to be the fixed fraction 6/52 of the whole dynamic range of the logged spectral estimate, i.e., the difference between the maximum and minimum value of the log spectrum. The fraction 6/52 was arrived at through empirical considerations. But in our development, we allow τ_f to be user-determined, allowing for additional flexibility.

A peak functional that corresponds to the Visual Significance approach of Soukup and Findley (1999) can be defined via

$$\Theta_{\theta,\delta}[f] = \min\{f(\theta) - f(\theta - \delta), f(\theta) - f(\theta + \delta)\}. \tag{3}$$

We will refer to $\Theta_{\theta,\delta}[\log f]$ as the VS functional at θ, or simply VS for short. Note that $\Theta_{\theta,\delta}$ is *not* a linear functional of spectra, due to the minimum in its definition. The criterion of VS states that a visually significant peak exists at frequency θ if $\Theta_{\theta,\delta}[\log f]$ exceeds τ (which is specified by the practitioner, and may depend on the dynamic range). In the specific settings of Soukup and Findley (1999), we would consider whether $\Theta_{j\pi/6,\pi/60}[\widehat{\log f}]$ (for $j = 1, 2, ..., 6$) exceeds τ_f, where τ_f is taken to be the fraction 6/52 times the dynamic range of the log spectrum. Of course, altering this τ_f to other values will naturally change the notion of peak – lowering τ_f makes our criterion less demanding.

We might study $\Theta_{\theta,\delta}[\log f]$ for various peak-shaped spectral densities f to get a reasonable notion of how to set the value of τ_f. Thus, it seems useful to examine this quantity for f given by (2), first letting θ equal any of the seasonal frequencies and $\delta = \pi/60$. Peak strength for the first and fifth seasonal frequencies is identical for this type of spectrum, and the same holds for the second and fourth seasonal frequencies. Figure 4 exhibits the log of peak strength, referred to as $\log \tau$, as a function of $\rho \in [.9, 1)$.

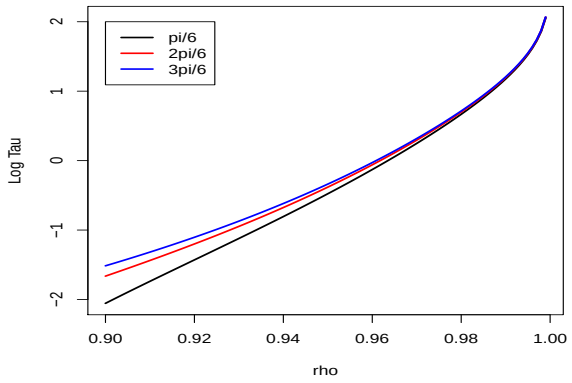


Figure 4: Plots of the logarithm of peak strength, as a function of persistence $\rho \in [.9, 1)$. Different curves correspond to the first, second, and third seasonal frequencies.

It is noteworthy that the log of peak strength, in this case, is an almost linear function of persistence ρ. This is noteworthy, because ρ might be taken as a proxy for the concept of peakedness; however, a defect of examining the magnitude of the roots of an AR polynomial that has been fitted to a time series, is that peak location does not exactly correspond to the angular portion of the roots; recall from (2) that the actual peak occurs at $\arccos[(1+\rho^2)\cos(\theta)/(2\rho)]$. When considering higher order AR processes, the log spectrum is the sum of logs of the squared magnitudes of the factors related to the zeroes of the AR polynomial. Contributions of the individual summands to the peaks of the AR spectrum are generally not identiable from the peak and hence the AR roots are not meaningful as measures of peak location and strength. On the other hand, the peak strength measure (3) is still meaningful when (2) fails to be true. For example, if the spectrum has shape

$$f(\lambda) = \sigma^2 \left(2 - 2\cos(\theta - \lambda)\right) \left(2 - 2\cos(\theta + \lambda)\right)^{-d}$$

for $d \in (0, 1/2)$, which corresponds to a long memory specification of seasonality (see discussion in Holan and McElroy (2012)), then peak strength $\Theta_{\theta,\delta}[\log f]$ becomes linear in d, i.e., peak strength is linear in the parameter d, whereas in the AR(2) case the log of peak strength is almost linear in ρ. It can be observed that for the adjusted series' peaks in Figures 1, 2, and 3 the peak strength is actually negative, because of the slender trough at θ. In practice, this is not a concern. We only care

to define a sensible range of estimands τ_f for peak strength, which can utilize the numbers displayed in Figure 4. In the case of an actual series that has been seasonally adjusted with SNR that is too high, the spectral estimate will not exhibit a trough unless there is a great deal of resolution in the grid, because spectral estimates involve smoothing over nearby frequencies. Hence, spectral estimates in the case of high SNR will tend to just show a peak, whereas when the SNR is too low, the spectral estimates will exhibit a trough.

2.3 A parametric Class of Spectral Densities

The discussion using the AR(2) spectrum and the long-memory spectrum shows that understanding what constitutes a peak and formulating a null hypothesis of absence of peak is a non-trivial exercise. Visual perception of a peak can be subjective and quantifying the peak measure that accurately reflects broad consensus about a peak can be challenging. The main difficulty is to derive a single measure that can adequately reflect the degree of peakedness of a function at a given location. This is because peakedness of a function is intrinsically a feature that needs a multi-dimensional specification, including

1. descriptions of the height of the function at that location
2. the concept of a width of the peak
3. the height of the baseline from where the peak rises
4. the convexity/concavity of the peak.

Other attributes that may contribute toward a description of the peak include asymmetry of the peak and rapid changes to slope.

Ideally, to understand the contribution of any single attribute one could observe the change in the peak resulting from changing the value of that attribute while fixing the remaining attributes. Such an exercise is not possible in the commonly used spectral classes (e.g., AR(2)) where the attributes are inter-related through the parameterization of the class. As a consequence, for commonly used models there is no obvious candidate for an appropriate peak functional which combines the attributes into a single measure that behaves monotonically with respect to any qualitative change in the peak. However, it's possible to define a new class of spectral models that parameterizes the relevant attributes separately, such that a peak functional of the form (3) will accurately quantify the notion of a peak.

The features that we want to study are the locations of the peaks (which in the present context will be the seasonal frequencies), the height at each peak location (which will be the values of the spectral density at the peak locations) as well as the widths of the peaks (an interval around the peak frequency within which all the peak dynamics are concentrated), the base height from where

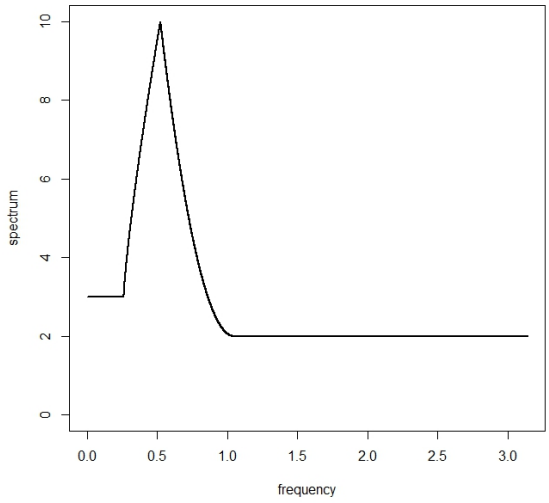


Figure 5: Triangular spectral density with peak at $\omega = \pi/6$.

the peak rises (the height of the spectral density at the end of the interval of frequencies depicting the width around a peak) and the curvature during the rise (whether mound-like or cusp-like). In addition we want the base heights and the curvatures to be different on either side of each peak to allow for asymmetry. A prototypical triangular shaped (hence the notation Tr) single peak function would be

$$Tr(\lambda|\theta, h, \delta_L, \delta_R, h_L, h_R, \alpha_L, \alpha_R) = \begin{cases} h_L, & 0 \le \lambda \le \theta - \delta_L, \\ h_L + (h - h_L)\left(\frac{\lambda - \theta + \delta_L}{\delta_L}\right)^{\alpha_L}, & \theta - \delta_L < \lambda \le \theta, \\ h_R + (h - h_R)\left(\frac{\theta + \delta_R - \lambda}{\delta_R}\right)^{\alpha_R}, & \theta < \lambda \le \theta + \delta_R, \\ h_R, & \theta + \delta_R < \lambda \le \pi. \end{cases} \tag{4}$$

The frequency θ will be the location of the peak and the height at the peak will be h. The parameters δ_L and δ_R control the width of the peak to the left and right, respectively. Similarly h_L and h_R are the base heights and α_L and α_R are the curvatures to the left and the right, respectively. A Triangular spectral density with a peak at the first seasonal frequency $\theta = \pi/6$ is shown in Figure 5. The associated parameters are $h = 10, h_l = 3, h_R = 2, \delta_L = \pi/12, \delta_R = \pi/6, \alpha_L = .8, \alpha_R = 2$.

Suppose we want to construct a spectral density with peaks at J target locations (e.g. seasonal frequencies) $0 < \theta_1 < \cdots < \theta_J < \pi$. For simplicity we are choosing the target frequencies in the open interval $(0, \pi)$. However, the function can be extended to include 0 and π. A spectral density with peaks at the target locations is then defined by

$$f(\lambda) = \sum_{j=1}^{J} Tr(\lambda|\theta_j, h_j, \delta_{Lj}, \delta_{Rj}, h_{Lj}, h_{Rj}, \alpha_{Lj}, \alpha_{Rj}). \tag{5}$$

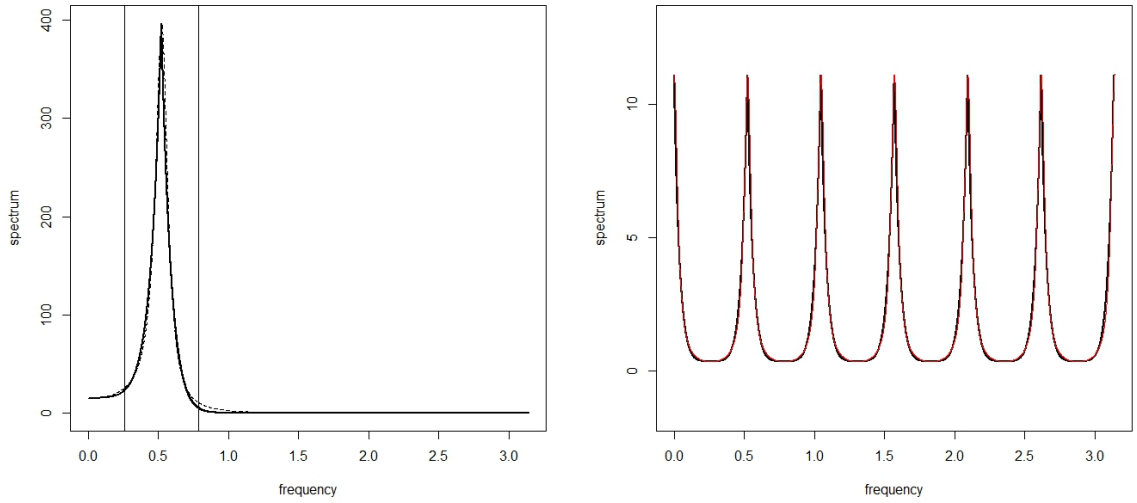
The heights and the base heights associated with the peaks in (5), due to superposition of several functions, are no longer the individual heights and base heights of the individual triangular components. However, in the case when the peak locations are well separated, i.e., $\theta_j + \delta_{Rj} < \theta_{j+1} - \delta_{L,j-1},\ j = 1, \ldots, (J-1)$, one could specify the heights and the base heights and obtain the parameters h, h_L and h_R from a set of constraints. If $b_1, \ldots, b_{J+1}$ are the desired base heights in the $(J+1)$ gaps between the peaks, and $p_1, \ldots, p_J$ are the desired heights at the peaks, then the parameters $(h_j, h_{Lj}, h_{Rj}), j = 1, \ldots, J$, can be obtained by setting $h_j = p_j$ for $j = 1, \ldots, (J-1)$, and by solving

$$\begin{aligned} \sum_{k=j}^{J} h_{Lk} + \sum_{k=1}^{j-1} h_{Rk} &= b_j,\ j = 1, \ldots, (J+1), \\ h_j + \sum_{k=2}^{J} h_{Lk} + \sum_{k=1}^{j-1} h_{Rk}, &= p_j,\ j = 1, \ldots, J. \end{aligned} \tag{6}$$

Thus, given J, i.e, the number of single peak Triangular components, the spectral density (5) will be parameterized by $\underline{\theta} = (\theta_1, \ldots, \theta_J)'$, $\underline{p} = (p_1, \ldots, p_J)'$, $\underline{b} = (b_1, \ldots, b_{J+1})'$, $\underline{\delta}_L = (\delta_{L1}, \ldots, \delta_{LJ})'$, $\underline{\delta}_R = (\delta_{R1}, \ldots, \delta_{RJ})'$, $\underline{\alpha}_L = (\alpha_{L1}, \ldots, \alpha_{LJ})'$, $\underline{\alpha}_R = (\alpha_{R1}, \ldots, \alpha_{RJ})'$, where the associated h, h_L and h_R parameters are solved using (6). When it is needed in the context, we will denote the spectral density (5) as $f(\lambda | J, \underline{\theta}, \underline{b}, \underline{\delta}_L, \underline{\delta}_R, \underline{\alpha}_L, \underline{\alpha}_R)$.

The class of functions (5) is quite flexible in terms of capturing the behavior of a peak. Figure 6(a) shows an approximation to the AR(2) spectrum in (1) (with parameters $\theta = \pi/6$ and $\rho = 0.9$) in terms of a Triangular spectrum. The two vertical lines show the frequency band $\pi/6 \pm \pi/12$ which provides a wide enough neighborhood between two consecutive seasonal frequencies. The Triangular spectrum seems to adequately approximate the peak feature of the AR(2) spectrum. Similarly, Figure 6(b) provides approximation to an SAR(1) spectrum (with parameter $\phi = 0.9$) using a Triangular spectral density with six peaks at the six seasonal frequencies. Again the degree of approximation is more than adequate for describing all relevant features of the spectral peaks of the SAR(1) spectrum using those from the Triangular spectral approximation.

The main purpose for introducing the new class of spectral densities is to study the contribution of individual characteristics such as peak height, peak width, and convexity towards the presence of seasonality. A peak at a seasonal frequency is a manifestation of periodic components of the associated frequency in the autocorrelation function of the time series. Ideally we would like to catalog the relationship between periodicity in the autocorrelation function and different shapes of a spectral function with a peak at a given location. Thus, it is imperative that we obtain an explicit expression for the autocorrelation function of the Triangular spectrum, which in turn can be used to study how presence of seasonality in the autocorrelation is affected by changes in any particular peak characteristics. The following result gives the autocorrelation function of the Triangular spectrum.



(a) AR(2) (b) SAR(1)

Figure 6: Triangular spectral density with peak at $\theta = \pi/6$.

Proposition 1. *Suppose* $f(\lambda)$ *given in* (5) *is the spectral density of a stationary process* $\{X_t\}$. *Then the autocorrelation function of the process at lag* k *is given by*

$$\gamma(k) = \sum_{j=1}^{J} \gamma_{Tr}(k|\theta_j, h_j, h_{Lj}, h_{Rj}, \delta_{Lj}, \delta_{Rj}, \alpha_{Lj}, \alpha_{Rj})$$

and

$$\begin{aligned}\gamma_{Tr} \;=\; & \pi^{-1}\{\theta(h_L - h_R)sinc(\theta k) + \pi h_R 1_{k=0} + \\ & \frac{\delta_L(h-h_L)}{2(\alpha_L+1)}\Big[e^{i(\theta-\delta_L)k}{}_1F_1(\alpha_L+1, \alpha_L+2, i\delta_L k) + e^{-i(\theta-\delta_L)k}{}_1F_1(\alpha_L+1, \alpha_L+2, -i\delta_L k)\Big] + \\ & \frac{\delta_R(h-h_R)}{2(\alpha_R+1)}\Big[e^{-i(\theta+\delta_R)k}{}_1F_1(\alpha_R+1, \alpha_R+2, i\delta_R k) + e^{i(\theta+\delta_R)k}{}_1F_1(\alpha_R+1, \alpha_R+2, -i\delta_R k)\Big]\},\end{aligned}$$

where $\gamma_{Tr} = \gamma_{Tr}(k|\theta, h, h_L, h_R, \delta_L, \delta_R, \alpha_L, \alpha_R)$, 1_A *is the indicator of the set* A, *the parameters* h, h_L *and* h_R *are solved from* $\underline{p}$ *and* $\underline{b}$ *using* (6), ${}_1F_1(\cdot,\cdot,\cdot)$ *is the confluent hypergeometric function defined in Gradshteyn and Ryzhik (1994, 9.210) and for* $x = 0$ *is it understood* $sinc(x) = \frac{sin(x)}{x} = 1$.

We use the proposed class as data generating process (in addition to AR(2) processes with single peak) in the simulation section to study the effect of different features of spectral peaks on the peak testing methodology proposed in the next section.

3 Peak Detection

Our main goal is to develop a hypothesis testing framework for the peak detection problem. To do so we must clearly describe the null and the alternative hypotheses in terms of parameters/functionals of the spectral density. Intuitively, if there is a peak functional $P[f]$ which captures the peak strength, absence of a peak would be quantified by the event that $P[f]$ fails to exceed some threshold. Because we are particularly interested in evaluating the VS framework within the paradigm of statistical significance, we closely adhere to the concepts and notions developed by Soukup and Findley (1999). Thus, we choose the peak functional as the VS functional, i.e., $P[f] = \Theta_{\theta,\delta}[\log f]$ (we define the VS functional in the log scale, which is more suitable for developing pivotal quantities).

Choosing an appropriate threshold is a critical component of the testing framework. It is intuitively appealing to let the threshold depend on the underlying spectral density, thereby making assessment of peak strength on a relative scale rather than using an absolute value. Henceforth we denote the threshold by τ_f to denote the dependence on f. In Soukup and Findley (1999) where the VS was introduced as a diagnostic, the sample version of the threshold was chosen to be $6/52$ of the estimated dynamic range of the spectral density. In the numerical analysis section we investigate different choices for the threshold.

The statistical peak testing problem as posed in the VS framework is not straight-forward. The main complication stems from the very definition of a peak, which makes it hard to clearly define the null set. The null hypothesis must include all cases which the user would generally classify as a non-peak. This includes spectral features that resemble half-peaks, where on one side of the target frequency the measure fails to qualify as peaked, even though on the other side the measure may indicate a strongly peaked feature. A pictorial illustration of such a case is given in Figure 7.

For convenience, we denote the right and the left-side peak measures included in the VS functional as $\Theta^L_{\theta,\delta}[\log f] = \log f(\theta) - \log f(\theta - \delta)$ and $\Theta^R_{\theta,\delta}[\log f] = \log f(\theta) - \log f(\theta + \delta)$. Thus, $\Theta_{\theta,\delta}[\log f] = \min\{\Theta^L_{\theta,\delta}[\log f], \Theta^R_{\theta,\delta}[\log f]\}$. Recognizing that an insignificant peak is one where at least one side is not significantly peaked, we obtain the following null and alternative hypotheses:

$$H^0_\theta : \Theta^L_{\theta,\delta}[\log f] \le \tau_f \text{ OR } \Theta^R_{\theta,\delta}[\log f] \le \tau_f \quad \text{vs} \quad H^a_\theta : \Theta^L_{\theta,\delta}[\log f] > \tau_f \text{ AND } \Theta^R_{\theta,\delta}[\log f] > \tau_f. \tag{7}$$

Intuitively, a critical region for such a test (with Type I error approximately equal to α) should look like

$$\mathcal{C}_\alpha = \{\hat{f} : \min\{\Theta^L_{\theta,\delta}[\log \hat{f}], \Theta^R_{\theta,\delta}[\log \hat{f}]\} > c_\alpha + \tau_f\}, \tag{8}$$

where $\hat{f}$ is some spectral estimator based on a time series sample of length n. Also, c_α is a constant chosen based on the joint distribution of $\Theta^L_{\theta,\delta}[\log \hat{f}]$ and $\Theta^R_{\theta,\delta}[\log \hat{f}]$, so that the size of the test (up to large sample approximation) is

$$\sup_{f:\, \Theta_{\theta,\delta}[\log f] \le \tau_f} P(\mathcal{C}_\alpha) \simeq \alpha,$$

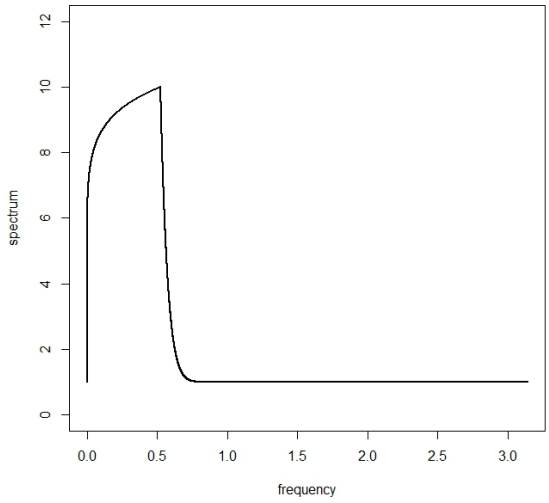


Figure 7: A spectral density where the right side VS at the seasonal frequency is high but the left side VS is low, and thus the dynamics around seasonal frequency $\pi/6$ will be classified as nonpeak.

where $\simeq$ stands for "asymptotically equal to". Because $P(\mathcal{C}_\alpha)$ is increasing in $\Theta_{\theta,\delta}[\log f]$, the supremum is achieved for f such that $\Theta_{\theta,\delta}[\log f] = \tau_f$. The test that rejects the hypothesis of no peak based on a rejection region of the form (8) will be called the VS_{new} test as opposed to the existing Visual Significance approach that uses the VS funnctional as a diagnostic tool. Below we consider conditions for which $\sup_{f:\,\Theta_{\theta,\delta}[\log f]=\tau_f} P(\mathcal{C}_\alpha) \simeq \alpha$. If the threshold τ_f is not known, a suitable sample estimate $\widehat{\tau_f}$ could be substituted. To further develop the hypothesis testing framework we need to describe the following:

1. What is the spectral estimator $\hat{f}$?
2. What is the relevant distribution theory for $\Theta_{\theta,\delta}[\log \hat{f}]$ that would provide appropriate pivotal quantities, which can then be inverted to obtain the α-critical region $\mathcal{C}_\alpha$?
3. How should one choose τ_f (or $\widehat{\tau_f}$)?

We define the spectral estimator $\hat{f}$ in the next section, and develop the relevant distribution theory. We also discuss the choice of τ_f and $\widehat{\tau_f}$ in the numerical analysis section. Before we move forward it is important to look at the structure of the peak testing problem in more detail. Typically, the distribution theory of the spectral estimator $\hat{f}$ would yield asymptotic pivotal quantities $(X_L, X_R) \equiv (\Theta^L_{\theta,\delta}[\log \hat{f}] - \Theta^L_{\theta,\delta}[\log f], \Theta^R_{\theta,\delta}[\log \hat{f}] - \Theta^R_{\theta,\delta}[\log f])$. For the peak testing problem, the marginal distributions of X_L and X_R are the same.

Proposition 2. *Let X_L and X_R be the aysmptotic pivots constructed from the left and right peak functionals. Let c_α be the upper α-percentile of the asymptotic distribution of X_L (or X_R). Then the critical region* (8) *is an asymptotically size α test for the hypothesis* (7)*.*

Proof. We will show that $P(C_\alpha) \lesssim \alpha$ for all f with $\Theta_{\theta,\delta}[\log f] = \tau_f$, and exhibit an f that attains this bound as well. Consider a value of the peak functionals on the boundary of the null hypothesis, i.e., consider f such that $\Theta_{\theta,\delta}[\log f] = \tau_f$. Without loss of generality let $\tau_f = \Theta^L_{\theta,\delta}[\log f] \leq \Theta^R_{\theta,\delta}[\log f]$. Then

$$\begin{aligned} P(\min\{\Theta^L_{\theta,\delta}[\log \hat{f}], \Theta^R_{\theta,\delta}[\log \hat{f}]\} > c_\alpha + \tau_f) &= P(\min\{X_L, X_R + \Theta^R_{\theta,\delta}[\log f] - \Theta^L_{\theta,\delta}[\log f]\} > c_\alpha) \\ &\leq P(X_L > c_\alpha) \simeq \alpha. \end{aligned}$$

Thus, $P(\mathcal{C}_\alpha : \Theta_{\theta,\delta}[\log f] = \tau_f) \lesssim \alpha$. Now consider the sequence of random variables $X_\mu = \min\{X_L, X_R + \mu - \tau_f\}$ which converges almost surely and hence weakly to X_L as $\mu \to \infty$. Therefore, with $\mu = \Theta^R_{\theta,\delta}[\log f]$,

$$\begin{aligned} P(\min\{\Theta^L_{\theta,\delta}[\log \hat{f}], \Theta^R_{\theta,\delta}[\log \hat{f}]\} > c_\alpha + \tau_f) &= P(\min\{X^L, X^R + \Theta^R_{\theta,\delta}[\log f] - \tau_f\} > c_\alpha) \\ &\to P(X_L > c_\alpha) \simeq \alpha, \end{aligned}$$

as $\Theta^R_{\theta,\delta}[\log f] \to \infty$. Thus, $\sup_{f:\, \Theta_{\theta,\delta}[\log f] \leq \tau_f} P(\mathcal{C}_\alpha) \simeq \alpha$ as well. □

Remark 1. Proposition 2 shows that in order to bound Type I error, the marginal pivot distribution should be used; this notion is further explored via simulation in Section 5.4.

The hypothesis about the peak functional defined as the minimum of two quantities can be rewritten in terms of the bivariate measure involving the left and the right side functionals separately. Then the hypothesis testing framework for testing (7) is analogous to one sided tests for linear inequality constraints; see Sasabuchi (1980), Berger (1989), and Liu and Berger (1995). As observed in Sasabuchi (1980) and Berger (1989), more powerful tests can be devised using the bivariate quantity than those defined simply using the minimum of the two random variables. This will be particularly true when the asymptotic distribution of the pivot is normal, e.g. in the case of AR-spectral estimator. However, since the main focus of this article is to evaluate the VS functional within a hypothesis testing paradigm, we will not elaborate on bivariate tests any further.

4 Statistical Estimation of Spectra

We consider tapered acf estimators defined as linear weighted combination of the first M sample autocovariances, with the weights determined by a specified taper. The quantity M will be referred to as the bandwidth length. Depending on how fast the bandwidth length grows relative to sample size, asympotically speaking, three different convergences are obtained. These are the cases of fixed bandwidth length, small bandwidth fraction, and fixed bandwidth fraction respectively. Following recent work on fixed bandwidth fraction asymptotics that established possibly superior finite sample performance as compared to the other bandwidth regimes in a variety of time series testing

problems, we concentrate on this case for the rest of the paper. Note that the spectral theory here is developed for stationary series, and hence presumes transformations and/or differencing has tamed any native non-stationarity.

4.1 Tapered Acf Spectrum Estimator

For background on the motivation and analysis of the tapered acf estimator of the spectrum, see McElroy and Politis (2014). Let $\langle g \rangle = (2\pi)^{-1} \int_{-\pi}^{\pi} g(\omega) d\omega$. Then the true autocovariance at lag k is $\widetilde{\gamma}_k = \langle \widetilde{f} c_k \rangle$, where $c_k(\theta) = \cos(\theta k)$. Supposing we have a sample $X_1, X_2, \cdots, X_n$, let us denote the periodogram via

$$\widehat{f}(\theta) = n^{-1} \left| \sum_{t=1}^{n} X_t e^{-i\theta t} \right|^2,$$

from which it follows that the sample acf is $\widetilde{\gamma}_k = \langle \widehat{f} c_k \rangle$. The periodogram is too ragged to consistently estimate $\widetilde{f}$, and is therefore smoothed out by tapering. A taper is an even function Λ that places more weight towards low lag sample acfs. The tapered acf estimator is defined as

$$\widehat{f}_{\Lambda,M}(\theta) = \sum_h \Lambda(h/M) \widehat{\gamma}_h c_h(\theta).$$

Here M is the bandwidth length, chosen by the practitioner. Often the taper is compactly supported on $[-1, 1]$, so that no lags greater than M are considered in the estimator. As a basic example, consider the truncation taper $\Lambda = 1_{[-1,1]}$, which weights all sample acf equally but only uses the first M of them. If we set $M = n$, we recover the periodogram:

$$\widehat{f} = \widehat{f}_{1_{[-1,1]},n}.$$

Another popular choice is the Bartlett taper $\Lambda_{BART}(x) = (1 - |x|) 1_{[-1,1]}(x)$; such a taper guarantees that the spectral estimate is non-negative, which is a desirable property when it is necessary to compute the logarithm of spectra.

4.2 Fixed Bandwidth Fraction Tapered Acf Estimator

The fixed bandwidth fraction paradigm considers that $M/n \to b \in (0, 1]$, a fixed fraction. This b is call the bandwidth fraction. For developing the fixed bandwidth fraction theory some data assumptions are needed. We present two main results below: first, the joint limiting behavior of the data's sine and cosine transforms as a functional limit theorem; second, the joint limiting behavior of the tapered acf spectral estimates. Excluding the long memory cases of McElroy and Politis (2014) for simplicity, we focus on three possible data assumptions:

- *Process* P1. $\{X_t\}$ is a linear process: $X_t = \sum_j \psi_j \epsilon_{t-j}$ with $\{\psi_j\}$ square summable and $\{\epsilon_t\}$ iid with finite variance.

- *Process* P2.. $X_t = g(Z_t)$ for each t, where g is a function in $\mathbb{L}^2(\mathbb{R}, e^{-x^2/2})$ of Hermite rank τ, and $\{Z_t\}$ is a Gaussian process.

- *Process* P3. $\{X_t\}$ is a strictly stationary process whose kth order cumulants exist and are summable over its k indices, for all $k \geq 1$.

Assumption P1 is standard fare, whereas P2 allows us to consider some non-linear processes. Assumption P3 allows for non-Gaussian processes, but with short-range dependence. These are viewed as unverifiable conditions that instead indicate the range of processes for which we can expect validity.

We are interested in establishing joint Discrete Fourier Transform (DFT) results over a sequence of frequencies near a finite set of target frequencies. Let the target frequencies be $\{\theta_1^0, \ldots, \theta_J^0\}$. Let $\{\theta_{1n}, \ldots, \theta_{Jn}\}$ be a sequence of frequencies over which we will evaluate the large sample properties of the DFTs. For any typical frequency in the sequence and the corresponding target, we use the notation $[\theta_{jn}, \theta_j^0]$ to denote the leading n^{-1} term in the frequency as

$$\theta_{jn} = \theta_j^0 + n^{-1}[\theta_{jn}, \theta_j^0] + o(n^{-1}).$$

Thus, the θ_j^0 are fixed target frequencies being approached at a certain rate over the sequence of frequencies θ_{jn}. For notational convenience we will drop the sample size dependence from the frequencies and simply denote them as $(\theta_1, \ldots, \theta_J)$. The real and the imaginary parts of the DFT at a frequency θ_j are the cosine and the sine transforms at that frequency and they are defined as

$$\begin{aligned} S_n^c(\theta_j) &= \sum_{t=1}^{n} X_t \cos(\theta_j t), \\ S_n^s(\theta_j) &= \sum_{t=1}^{n} X_t \sin(\theta_j t). \end{aligned} \tag{9}$$

The distribution theory used in the proposed testing procedure for the peak functional will depend on the asymptotic distribution of the sine and cosine transforms at relevant frequencies. When θ_j^0 is equal to zero or π, the asymptotic results are a bit different from the other cases where $\theta_j^0 \in (0, \pi)$. We first give an expression for the limiting covariances

$$V_{cc}(\theta_j, \theta_k) := cov(S_n^c(\theta_j), S_n^c(\theta_k)) \tag{10}$$

$$V_{cs}(\theta_j, \theta_k) := cov(S_n^c(\theta_j), S_n^s(\theta_k)) \tag{11}$$

$$V_{ss}(\theta_j, \theta_k) := cov(S_n^s(\theta_j), S_n^s(\theta_k)) \tag{12}$$

for a typical pair of frequencies (θ_j, θ_k). Also, for pairs of frequencies along with corresponding target frequencies, define the sets

$$\begin{aligned} \mathcal{A} &= \{(\theta_j, \theta_k) \in [0, \pi] \times [0, \pi] : \theta_j^0 + \theta_k^0 = 0, 2\pi \text{ (modulo } 2\pi)\} \\ \mathcal{B} &= \{(\theta_j, \theta_k) \in [0, \pi] \times [0, \pi] : \theta_j^0 - \theta_k^0 = 0, 2\pi \text{ (modulo } 2\pi)\}. \end{aligned} \tag{13}$$

Proposition 3. *Let $(S_n^c(\theta_j), S_n^s(\theta_k)$ be the cosine and sine transformations* (9) *for frequencies θ_j and θ_k, respectively, and let $V_{cc}(\theta_j,\theta_k), V_{cs}(\theta_j,\theta_k)$ and $V_{ss}(\theta_j,\theta_k)$ denote the covariances as defined in* (12)*. Then*

$$\begin{aligned} V_{cc}(\theta_j,\theta_k) &= n\underline{V}_{cc}(\theta_j,\theta_k) + o(n), \\ V_{cs}(\theta_j,\theta_k) &= n\underline{V}_{cs}(\theta_j,\theta_k) + o(n), \\ V_{ss}(\theta_j,\theta_k) &= n\underline{V}_{ss}(\theta_j,\theta_k) + o(n), \end{aligned}$$

where

$$\begin{aligned} \underline{V}_{cc}(\theta_j,\theta_k) &= \frac{1}{2}\left(\frac{f(\theta_j^0)+f(\theta_k^0)}{2}\right)\left(\frac{\sin([\theta_j+\theta_k,\theta_j^0+\theta_k^0])}{[\theta_j+\theta_k,\theta_j^0+\theta_k^0]}1_{\mathcal{A}} + \frac{\sin([\theta_j-\theta_k,\theta_j^0-\theta_k^0])}{[\theta_j-\theta_k,\theta_j^0-\theta_k^0]}1_{\mathcal{B}}\right) \\ &\quad -(g(\theta_j^0)+g(\theta_k^0))\left(\frac{1-\cos([\theta_j+\theta_k,\theta_j^0+\theta_k^0])}{[\theta_j+\theta_k,\theta_j^0+\theta_k^0]}1_{\mathcal{A}}\right) \\ &\quad -(g(\theta_j^0)-g(\theta_k^0))\left(\frac{1-\cos([\theta_j-\theta_k,\theta_j^0-\theta_k^0])}{[\theta_j-\theta_k,\theta_j^0-\theta_k^0]}1_{\mathcal{B}}\right), \\ \underline{V}_{cs}(\theta_j,\theta_k) &= \frac{1}{2}\left(\frac{f(\theta_j^0)+f(\theta_k^0)}{2}\right)\left(\frac{1-\cos([\theta_j+\theta_k,\theta_j^0+\theta_k^0])}{[\theta_j+\theta_k,\theta_j^0+\theta_k^0]}1_{\mathcal{A}} - \frac{1-\cos([\theta_j-\theta_k,\theta_j^0-\theta_k^0])}{[\theta_j-\theta_k,\theta_j^0-\theta_k^0]}1_{\mathcal{B}}\right) \\ &\quad +(g(\theta_j^0)+g(\theta_k^0))\left(\frac{\sin([\theta_j+\theta_k,\theta_j^0+\theta_k^0])}{[\theta_j+\theta_k,\theta_j^0+\theta_k^0]}1_{\mathcal{A}}\right) \\ &\quad -(g(\theta_j^0)-g(\theta_k^0))\left(\frac{\sin([\theta_j-\theta_k,\theta_j^0-\theta_k^0])}{[\theta_j-\theta_k,\theta_j^0-\theta_k^0]}1_{\mathcal{B}}\right), \\ \underline{V}_{ss}(\theta_j,\theta_k) &= \frac{1}{2}\left(\frac{f(\theta_j^0)+f(\theta_k^0)}{2}\right)\left(\frac{\sin([\theta_j+\theta_k,\theta_j^0+\theta_k^0])}{[\theta_j+\theta_k,\theta_j^0+\theta_k^0]}1_{\mathcal{A}} - \frac{\sin([\theta_j-\theta_k,\theta_j^0-\theta_k^0])}{[\theta_j-\theta_k,\theta_j^0-\theta_k^0]}1_{\mathcal{B}}\right) \\ &\quad +(g(\theta_j^0)+g(\theta_k^0))\left(\frac{1-\cos([\theta_j+\theta_k,\theta_j^0+\theta_k^0])}{[\theta_j+\theta_k,\theta_j^0+\theta_k^0]}1_{\mathcal{A}}\right) \\ &\quad -(g(\theta_j^0)-g(\theta_k^0))\left(\frac{1-\cos([\theta_j-\theta_k,\theta_j^0-\theta_k^0])}{[\theta_j-\theta_k,\theta_j^0-\theta_k^0]}1_{\mathcal{B}}\right), \end{aligned}$$

$g(\theta_j) = \sum_{h=1}^{\infty}\gamma_h \sin(\theta_j h)$ and the sets $\mathcal{A}$ and $\mathcal{B}$ are defined in (13)*.*

Proposition 3 allows us to write a joint functional limit theorem for the sine and cosine transformations at the frequencies $\{\theta_1,\ldots,\theta_J\}$. Theorem 1 gives the asymptotic distribution of the sine and cosine transforms at the given sequence of frequencies. Define stochastic processes from the sine and cosine transforms via $S^s_{[rn]}$ and $S^c_{[rn]}$ as a function of $r \in (0,1)$, and write $\xi^s_{[\cdot n]}$ and $\xi^c_{[\cdot n]}$ for the linearly interpolated versions. Our results below do not consider mean-centering, because in applications the time series have already been regression-adjusted and differenced to remove nonstationarity, so that typically the mean is zero. Extensions of the theorems below to a non-zero mean are possible, as outlined in McElroy and Politis (2014), but there is no impact except at

frequency zero anyways, and we have no interest in peak detection at frequency zero for purposes of assessing seasonal adjustment. For this reason, we restrict the target frequencies to $(0, \pi)$.

Theorem 1. *Let $\{X_t\}$ be a mean zero covariance stationary time series such that $\mathbb{E}[|X_t|^{2+\delta}] < \infty$ for some $\delta > 0$, and assume that $\mathbb{E}[|S_n^s(\theta)|^{2+\delta}]$ and $\mathbb{E}[|S_n^c(\theta)|^{2+\delta}]$ are $O(n^{1+\delta/2})$ uniformly in θ. Suppose that condition P1, P2, or P3 hold. Then for the sequence of frequencies $\{\theta_1, \ldots, \theta_J\}$ with target frequencies all in $(0, \pi)$, as $n \to \infty$*

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \xi^c_{\lfloor rn \rfloor}(\theta_1) \\ \xi^s_{\lfloor rn \rfloor}(\theta_1) \\ \vdots \\ \xi^c_{\lfloor rn \rfloor}(\theta_J) \\ \xi^s_{\lfloor rn \rfloor}(\theta_J) \end{pmatrix} \Longrightarrow \underline{V}^{1/2} \begin{pmatrix} B^c_1(r) \\ B^s_1(r) \\ \vdots \\ B^c_J(r) \\ B^s_J(r) \end{pmatrix} := \begin{pmatrix} A^c_1(r) \\ A^s_1(r) \\ \vdots \\ A^c_J(r) \\ A^s_J(r) \end{pmatrix},$$

where each process B^c_j and B^s_j for $1 \leq j \leq J$ is an independent Brownian motion. The covariance matrix of the vector $A(r)$ process for any fixed r is $\underline{V}$, with the covariance of $A^x_j(r)$ and $A^y_k(r)$ given by $\underline{V}_{xy}(\theta_j, \theta_k)$ in Proposition 3 for x, y denoting either c or s.

Using the functional limit theorem in Theorem 1, it is possible to write down the joint distribution of the tapered acf spectral estimator $\widehat{f}_{\Lambda,M}$ at finitely many frequencies, under a fixed bandwidth fraction assumption. For the next result, we allow Λ to be flat-top (i.e., there is a $c \in [0, 1)$ such that $\Lambda(x)$ is constant for $|x| \leq c$) and is piecewise twice continuously differentiable.

Theorem 2. *Let $\{X_t\}$ be a mean zero covariance stationary time series such that $\mathbb{E}[|X_t|^{2+\delta}] < \infty$ for some $\delta > 0$, and assume that $\mathbb{E}[|S_n^s(\omega)|^{2+\delta}]$ and $\mathbb{E}[|S_n^c(\omega)|^{2+\delta}]$ are $O(n^{1+\delta/2})$ uniformly in ω. Suppose that condition P1, P2, or P3 hold. Then for the sequence of frequencies $\{\theta_1, \ldots, \theta_J\}$, as $n \to \infty$*

$$\left(\frac{\widehat{f}_b(\theta_1)}{f(\theta_1)}, \ldots, \frac{\widehat{f}_b(\theta_J)}{f(\theta_J)} \right) \Longrightarrow (S_{\theta_1}(b), \ldots, S_{\theta_J}(b)).$$

The limiting random vector $S_{\theta_j}(b)$ is defined as

$$\begin{aligned} S_{\theta_j}(b) \;=\; & \frac{-1}{b^2} \int\int_{cb<|r-s|<b} \ddot{\Lambda}(\frac{r-s}{b})(A^c_{\theta_j}(r)A^c_{\theta_j}(s) + A^s_{\theta_j}(r)A^s_{\theta_j}(s)drdr \\ & + \frac{2}{b}\dot{\Lambda}_-(1) \int_0^{1-b} (A^c_{\theta_j}(r)A^c_{\theta_j}(r+b) + A^s_{\theta_j}(r)A^s_{\theta_j}(r+b))dr \\ & - \frac{2}{b}\dot{\Lambda}_+(c) \int_0^{1-bc} (A^c_{\theta_j}(r)A^c_{\theta_j}(r+bc) + A^s_{\theta_j}(r)A^s_{\theta_j}(r+bc))dr \\ & + \frac{2}{b} \int_{1-b}^{1-bc} \dot{\Lambda}(\frac{1-r}{b})(A^c_{\theta_j}(r)A^c_{\theta_j}(1) + A^s_{\theta_j}(r)A^s_{\theta_j}(1))dr \\ & + \Lambda(0) \left(A^c_{\theta_j}(1)A^c_{\theta_j}(1) + A^s_{\theta_j}(1)A^s_{\theta_j}(1) \right), \end{aligned}$$

where the processes $A^c_{\theta_j}$ and $A^s_{\theta_j}$ are as defined in Theorem 1, $\dot{\Lambda}(x), \ddot{\Lambda}(x), \dot{\Lambda}_-(x)$ and $\dot{\Lambda}_+(x)$ are the first, second, left and right derivatives of Λ at x, respectively. In the case that there is a jump discontinuity in Λ at c, we must replace the third summand in the limit distribution by

$$2(\Lambda^+(c) - \Lambda^-(c))\ (A^c_{\theta_j}(1-bc)A^c_{\theta_j)}(1) + (A^s_{\theta_j}(1-bc)A^s_{\theta_j)}(1)).$$

Allowing for jump discontinuities at c means the results cover the truncation taper, which is important for application of the theory to other forms of estimators such as fixed model order AR estimators (an estimator not discussed here, but could be thought of as a special case of the tapered acf estimator using a representation of the AR estimator by Berk (1974)). For the specific application of peak testing, the finitely many frequencies $\theta_1, \ldots, \theta_J$ involved in the construction of any local measure (by local we mean that only features in the neighborhood of the target seasonal frequency are considered for the determination of a peak) of peak strength, e.g.(3), will have the property $\theta_j = \theta^0 + [\theta_j, \theta^0]$. Thus, for any two frequencies θ_j, θ_k, the associated variance covariance expressions for the sine and cosine transforms reduce to

$$\begin{aligned}
V_{cc}(\theta_j,\theta_j) = V_{ss}(\theta_j,\theta_j) &= \frac{1}{2}f(\theta^0), \\
V_{sc}(\theta_j,\theta_j) &= 0, \\
V_{cc}(\theta_j,\theta_k) = V_{ss}(\theta_j,\theta_j) &= \frac{1}{2}f(\theta^0)\,\frac{\sin([\theta_j-\theta_k, \theta_j^0-\theta_k^0])}{[\theta_j-\theta_k, \theta_j^0-\theta_k^0]}, \\
V_{sc}(\theta_k,\theta_j) = -V_{cs}(\theta_k,\theta_j) &= \frac{1}{2}f(\theta^0)\,\frac{1-\cos([\theta_j-\theta_k, \theta_j^0-\theta_k^0])}{[\theta_j-\theta_k, \theta_j^0-\theta_k^0]}, \\
V_{cc}(\theta_k,\theta_k) = V_{cc}(\theta_k,\theta_k) &= \frac{1}{2}f(\theta^0), \\
V_{sc}(\theta_k,\theta_k) &= 0,
\end{aligned}$$

5 Numerical Studies

In this section we investigate the statistical properties of the method. We first discuss the choice of τ_f in the test (8) (referred to as VS_{new} hereafter), via examination of the AR(2) process (1); second, we explore the size and power of the VS_{new} method through simulation studies, and secondly evaluate the method on Census Bureau time series.

5.1 Choice of τ_f

For Soukup and Findley (1999), τ_f was chosen as $\tau_f = \tau \cdot R_{\log f}$ where $R_{\log f}$ is the range of the log spectrum (necessarily assumed to be bounded) and τ is a fixed constant equal to 6/52. Thus, Soukup and Findley (1999) considered a peak functional estimate not visually significant if the peak functional was less than 6/52 of the range of the estimated log-spectrum. Their choice of τ was guided by the available resolution in the plotting device (i.e., the star plot). Here we want to

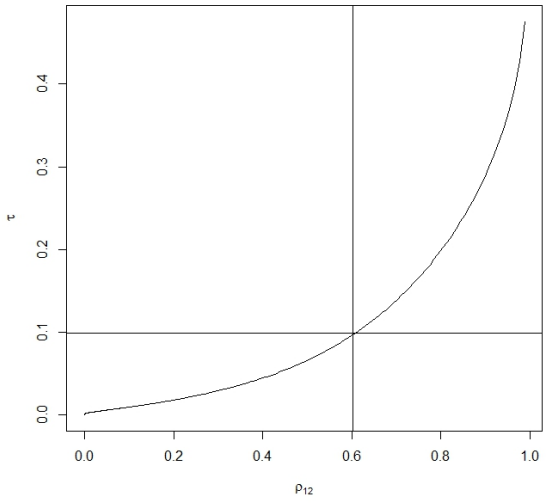


Figure 8: Seasonal lag (12) correlation versus the VS value τ for the AR(2) model

relate the choice of the null hypothesis value of τ_f to the absence of seasonality in the time series. While there is no consensus in the seasonal adjustment literature about a canonical quantification of seasonality, it seems reasonable to expect that the presence of a large dominant autocorrelation at the seasonal lag is sufficient for a time series to have seasonal behavior – see the discussion in Fase et al. (1973) and Hylleberg (1986). Let $\rho(k)$ denote the autocorrelation function (acf) associated with the spectrum. We define the peak in the acf at the seasonal lag L to be dominant if

RULE 1: $\rho(L) \geq$ THR and $\rho(L) \geq \rho(k)$ for $0.5L \leq k \leq 1.5L$,

where THR is some user specified threshold. The rule, simply put, means that there is a strong positive correlation at the seasonal lag, and in addition the seasonal lag correlation is bigger than any other autocorrelation at the neighboring lags.

For the AR(2) process (1) with a single peak at $\theta = \pi/6$, the value of $\tau = \Theta_{\theta,\delta}[\log f]/R_{\log f}$ is a monotone function of $\phi_2 = -\rho^2$, where $\theta = \pi/6$ and $\delta = \pi/60$. Similarly, $\rho(L)$ is also a monotone function of ϕ_2. Figure 5.1 shows the value of τ as a function of $\rho(12)$, since the seasonal lag of interest is 12 for a monthly series with peak at $\theta = \pi/6$.

For lower THR values (less than 0.2) the lag 12 autocorrelation is not dominant among the neighboring correlations ranging from lag 6 to lag 18. For the AR(2) process, a value of τ bigger than 0.1 satisfies **RULE 1** with THR ≈ 0.6. A value of $\tau = 0.067$ would correspond to THR equal to 0.5 for the AR(2) process. The value 0.1 is close to the value 6/52 used by Soukup and Findley (1999). Of course, the behavior of $\rho(k)$ for $|k - L| < L/2$ is not solely determined by the VS functional. As illustrated by our numerical results for other classes of spectral densities,

the attributes of the autocorrelation sequence around the seasonal lag are determined by several features of the spectral density at the seasonal peak, of which the VS functional seems to be one of the main features to consider. Based on the values obtained from the AR(2) process, we chose $\tau = 0.1$ for the remainder of the simulation experiment. A point to note is that the choice of τ is still a subjective issue, and the simulation results are presented only to demonstrate that the proposed test works for any given value of τ.

5.2 Choice of estimator of range of log-spectrum

As mentioned earlier, to develop the testing theory we assume that the true log-spectral density is bounded in absolute value (as assumption that is implicit in the definition of the VS functional). To use the proposed VS_{new} test we need to consistently estimate the range of the log-spectral density. To do that we use a higher order AR model to estimate the spectral density and compute the range of the log-spectral density based on the model estimate. The reason for choosing a parametric model is to reduce variability in the estimate. One could use nonparametric tapered estimators, but in our investigation we found that AR based estimator provided more stable range estimates, thereby reducing the variability in the rejection rate incurred from using a sample estimate of the range in the rejection rule. Of course, under a $p \to \infty$ paradigm, the AR(p) estimator will provide consistent estimate of the range (Tiao and Tsay, 1983). For implementation of the test we found that an AR(14) adequately captures the range of the log-spectral density for a wide range of models investigated.

5.3 AR(2) peak

Throughout the simulation study we concentrate on monthly series, and hence use lag 12 as the primary seasonal lag and $\theta = \pi/6$ as the dominant seasonal frequency. The AR(2) parameterization (1) provides a spectral density with a single peak at the first seasonal frequency ($\pi/6$) for monthly data. If the peak in the spectrum of a seasonally adjusted series is well-depicted by an AR(2) spectral peak, then we could evaluate the power of the VS_{new} test in detecting such spectral peaks using the AR(2) data generating process (DGP). For numerical investigation we generate series of length $n = 120$ corresponding to 10 year series, and also for $n = 600$ corresponding to 50 year series. The smaller sample size is similar to what is used in the X-12 ARIMA software (i.e., the final 8 years of data) in the context of visual significance, and the longer sample is used to investigate the convergence to asymptotic values. For 10 year series, the VS_{new} test failed to produce any reasonable power when $\delta = \pi/60$ was used in the VS definition. For $\delta = \pi/60$, the VS functional estimate unreliable since it is based on only two frequencies that falls within the δ width. In order to reduce variability, we also investigated wider δ equal to $\pi/15$.

Table 1 provides the power for the VS_{new} test for different values of the VS based on the

Bartlett spectral estimator. The true VS was fixed to be in a grid $\{0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$ and the parameter ρ of the AR(2) parameterization (1) was computed based on the VS value (in this case the VS is a monotone function of ρ and a simple grid search provided the associated ρ value for each VS value). The VS values equal to 0, 0.05, and 0.1 are in the null region, i.e. VS $\leq \tau = 0.1$. The AR(2) spectrum is fairly symmetric, providing similar values of $\Theta^L_{\theta,\delta}[\log f]$ and $\Theta^R_{\theta,\delta}[\log f]$, and thus the size of the VS_{new} test is expected to be conservative. We only report power values for $b \in [0.3, 0.8]$ because for values of b smaller than 0.3, the finite sample estimator over-smooths, and the VS_{new} test becomes more conservative. In contrast, for $b \geq 0.9$ there is increased variability in the estimator resulting in variability in the observed power.

The VS_{new} test is more conservative for smaller sample sizes; for $\delta = \pi/15$ the size of the test is about 2% for $n = 120$ whereas the size increases to about 3.3% (Table 2) for $n = 600$. Note that the difference of the size from the nominal $\alpha = 0.05$ is not as stark as one would expect if the left and right peak functionals where asymptotically independently distributed (which would result in a size of approximately $\alpha^2 = 0.0025$). Table 3 provides the power of the small-b test for $\delta = \pi/60$, thus using the same width for the VS as described in the Visual Significance approach of Soukup and Findley (1999). For $n = 600$ the test shows reasonable power against alternative VS values considered, with the test showing nearly 100% power for an alternative value of VS = 0.5. For $n = 600$ and $\delta = \pi/60$ there are about 10 Fourier frequencies that fall in either side of the target frequency within the δ range and hence provide reasonably stable local estimates. Thus, the performance of the VS_{new} test is sensitive to the relationship between the width δ and the sample size n.

Also, the investigation using an AR(2) process limits the shape and other features of the true spectral density, and hence it may not reveal sensitivity of the proposed method to the different aspects of a spectral peak. To address this issue we consider a new class of parametric models where the parameters are specifically chosen to depict individual features of spectral peaks, thereby allowing us to investigate the influence of each feature on the performance of the VS_{new} test separately.

5.4 Parametric Spectral Peaks

In this section we evaluate the performance of the VS_{new} test when the time series is generated based on the parametric model defined in Section 2.3. For simulation we first chose symmetric peaks where the width, the ratio of the peak height to the base height and the convexity were identical for the left and the right parts of the Triangular peaks. We chose the width of the Triangular peak to be $\pi/12$ making it wider than the width δ used in the definition of VS. This choice would ensure that VS is not simply based upon the ratio of the peak height and the base height which we chose to be 100. The convexity parameter was obtained by fixing the width, the ratio and the VS value.

	Bandwidth fraction b					
VS	.3	.4	.5	.6	.7	.8
0.000	0.002	0.004	0.005	0.006	0.006	0.007
0.050	0.002	0.004	0.004	0.005	0.007	0.007
0.100	0.008	0.009	0.011	0.013	0.015	0.016
0.150	0.034	0.042	0.050	0.053	0.055	0.058
0.200	0.059	0.078	0.092	0.097	0.098	0.101
0.300	0.328	0.367	0.391	0.397	0.390	0.380
0.400	0.716	0.742	0.764	0.761	0.752	0.744
0.500	0.926	0.933	0.938	0.939	0.935	0.931

Table 1: Power of proposed test based on small-b asymptotics for the AR(2) model for different values of the true visual significance measure. The values in bold correspond to type I error values since the associated VS values are all in the null region (≤ 0.1). The sample size is $n = 120$ and the width δ used in the VS measure is $\pi/15$.

As before, the VS_{new} test was very conservative for the smaller sample size ($n = 120$) when the VS width was $\delta = \pi/60$, and the case is not reported here. Tables 4-6 show the power of the VS_{new} test for the Triangular spectral model for different values of the bandwidth parameter b (again using Bartlett spectral estimator). The power of the proposed test is sensitive to the shape of the peak. Since the height of the peak and the base and the width of the peak are fixed, the VS value is a function of the convexity parameter with the higher VS values associated with more cusp-like peaks which has higher values for the convexity parameter. Thus, we can enhance the visual significance of the peak and hence the power of the VS_{new} test by altering the convexity while holding all other features fixed.

For the Triangular model, it is possible to look at the performance of the test when the peak is asymmetric and investigate the Type I error rate for null values where one side of the VS has a high value while the other side is within the null region. Without loss of generality, we let the left side VS value to be equal to the null boundary value of $\tau = 0.1$ and vary the right side VS value between 0.1 and 0.9. As expected (cf. Proposition 2), the Type I error rate approaches the nominal error rate α, equal to 5% for the current investigation, when the right VS value approaches 1 (or equivalently the right peak functional approaches infinity); see Table 7.

In contrast, if the upper α percentile of the distribution of the minimum of the left and right asymptotic pivots was used in the definition of the critical region, the type I error rates will be inflated for the asymmetric null case while giving nearly nominal rates for the symmetric null. Table 8 shows the type I error rates for a rejection rule that used the percentile of the distribution of the minimum of the pivots rather than the percentile of the distribution of the individual pivots. Thus the approach using the percentile of the distribution of the minimum of left and right pivots

	Bandwidth fraction b					
VS	.3	.4	.5	.6	.7	.8
0.000	0.013	0.013	0.014	0.016	0.015	0.016
0.050	0.010	0.012	0.013	0.013	0.011	0.013
0.100	0.032	0.031	0.031	0.033	0.030	0.032
0.150	0.114	0.107	0.099	0.095	0.086	0.085
0.200	0.314	0.273	0.246	0.226	0.196	0.185
0.300	0.851	0.786	0.733	0.690	0.636	0.612
0.400	0.995	0.989	0.982	0.974	0.960	0.947
0.500	1.000	1.000	1.000	0.999	0.999	0.999

Table 2: Power of proposed test based on small-b asymptotics for the AR(2) model for different values of the true visual significance measure. The values in bold correspond to type I error values since the associated VS values are all in the null region (≤ 0.1). The sample size is $n = 600$ and the width δ used in the VS measure is $\pi/15$.

would fail to control the Type I error rate at the nominal level α when asymmetric peaks are present. This underscores the importance of utilizing the (marginal) distribution of individual pivots, as established in Proposition 2.

5.5 Comparison with the existing Visual Significance approach

The existing Visual Significance appraoch uses a rule that declares the presence of a spectral peak if the VS functional value exceeds the 6/52 fraction of the observed dynamic range. A comparison of this rule – which we henceforth refer to as the "VS rule" – in the proposed testing framework would amount to a rejection rule (8) where the percentile c_α is set to zero. This would lead to inflated Type I error. Table 9 shows the Type I error rate and power for the VS rule under the AR(2) process. Similarly inflated Type I error rates are observed under the Triangular model as well, but for brevity we report on the AR(2). To make the results comparable with the proposed testing method, we have used the fraction 0.1 as the threshold for the VS rule, but the Type I error rate at 6/52 is very inflated as well.

6 Data Analysis

The application of VS in X-12-ARIMA pertains to the original data, to the RegARIMA model residuals, the SA series, and the estimated irregular. However, the raw data is typically nonstationary, so that tapered spectral estimators are not appropriate. For this reason, we focus only on stationary data.

More precisely, the tools of this paper properly apply to stationary data, and cannot be applied

	Bandwidth fraction b					
VS	.3	.4	.5	.6	.7	.8
0.000	0.008	0.009	0.011	0.012	0.014	0.013
0.050	0.001	0.003	0.003	0.004	0.004	0.005
0.100	0.009	0.015	0.018	0.018	0.019	0.019
0.150	0.044	0.049	0.053	0.056	0.060	0.054
0.200	0.128	0.156	0.160	0.158	0.165	0.149
0.300	0.597	0.647	0.648	0.631	0.625	0.588
0.400	0.918	0.932	0.927	0.918	0.916	0.897
0.500	0.988	0.989	0.989	0.987	0.986	0.983

Table 3: Power of the VS_{new} test based on small-b asymptotics for the AR(2) model for different values of the true visual significance measure. The values in bold correspond to type I error values since the associated VS values are all in the null region (≤ 0.1). The sample size is $n = 600$ and the width δ used in the VS measure is $\pi/60$.

to nonstationary data without a huge distortion to the asymptotic distribution theory. Whereas raw data will have trend and seasonal nonstationarities typically present, seasonal adjustments can be assumed to have only milder forms of seasonality present; trend-differencing the SA results in data that can be often supposed to be stationary. Our methods can then be applied to trend difference SA data, or to model residuals, or the estimated irregular.

Since the critical values of our procedure are based on a stationary null hypothesis, its application to raw data may merely result in very high rejection rate. Because the distributional properties for the nonstationary case are unknown, we do not make this application. Both the model residuals and irregular are stationary, and residual spectral peaks indicate potential problems with the adjustment.

6.1 Empirical analysis of series from multiple sectors

For data analysis, we consider monthly series from four different sectors: manufacturing (86 series), retail (6 series), wholesale (18 series) and residential construction (16 series). Each series begins in April 1995, and is 10 years long. We processed each series in X-12-ARIMA and noted the visually significant peaks detected by the VS diagnostics for the seasonally adjusted, irregular and residual series corresponding to each original series. In order to reduce the impact of human decisions on the adjustments, automatic modeling and outlier detection was used by X-12-ARIMA, along with trading day identification. We applied the VS_{new} test with the Bartlett spectral estimator to differenced seasonal adjustments, model residuals, and irregulars, reporting the p-values in each case and utilizing a null τ value of 0.1. This test was based upon the full data span of 20 years; but the default VS procedure only uses the final 8 years of data.

	Bandwidth fraction b					
VS	.3	.4	.5	.6	.7	.8
0.000	0.005	0.005	0.006	0.006	0.006	0.007
0.050	0.006	0.006	0.007	0.007	0.008	0.009
0.100	0.016	0.016	0.016	0.017	0.016	0.018
0.150	0.032	0.034	0.035	0.034	0.032	0.034
0.200	0.044	0.048	0.051	0.049	0.046	0.048
0.300	0.101	0.106	0.109	0.105	0.102	0.099
0.400	0.194	0.211	0.211	0.205	0.198	0.195
0.500	0.305	0.331	0.324	0.315	0.309	0.303

Table 4: Power of the VS_{new} test based on small-b asymptotics for the single peak Triangular model for different values of the true visual significance measure. The values in bold correspond to type I error values since the associated VS values are all in the null region (≤ 0.1). The sample size is $n = 120$ and the width δ used in the VS measure is $\pi/15$.

In terms of results, the majority of adjustments were adequate when the X-12-ARIMA VS criterion is used. There were 5 series (of 126) whose residuals contained a visually significant peak according to X-12-ARIMA. These series were "Other electronic component manufacturing" and "Communication equipment manufacturing" from the manufacturing sector, "Home under construction" for the northeast and south regions in the residential construction sector and the "Clothing store" series in the retails sales sector. Our proposed test found no significant peak in any of the series.

6.2 Diagnosis of residual seasonality after inadequate adjustment

We also used 16 retail series with suspected seasonal behavior and intentionally performed inadequate seasonal adjustment to check if the test could detect residual seasonality after the adjustment. The 16 series where "Retail and food services sales, total" (44000), "Electronics and appliance stores" (44300), "Computer and software stores" (44312), "Building materials and garden equipment and supplies dealers" (44400), "Grocery stores" (44510), "Clothing and clothing accessory stores" (44800), "Men's clothing stores" (44811), "Women's clothing stores" (44812), "Shoe stores" (44820), "Sporting goods, hobby, book, and music stores" (45100), "General merchandise stores" (45200), "Department stores -excluding leased departments" (45210), "Warehouse clubs and superstores" (45291), "Nonstore retailers" (45400), "Electronic shopping and mail-order houses" (45410) and "Food services and drinking places" (72200). The start and end dates for each series were January 1992 and December 2007, respectively. Thus the length of each series was n = 192. The series are suspected to have changing seasonality (based on the fact that X-12-ARIMA chose short filters for each of the series) and hence each calendar month's average seasonality for the last 10

	Bandwidth fraction b					
VS	.3	.4	.5	.6	.7	.8
0.000	0.001	0.002	0.001	0.002	0.001	0.002
0.050	0.005	0.006	0.007	0.008	0.008	0.008
0.100	0.013	0.015	0.017	0.016	0.014	0.016
0.150	0.037	0.034	0.034	0.034	0.033	0.034
0.200	0.080	0.074	0.069	0.071	0.062	0.062
0.300	0.285	0.234	0.201	0.192	0.170	0.161
0.400	0.564	0.476	0.412	0.376	0.331	0.312
0.500	0.802	0.723	0.657	0.612	0.549	0.522

Table 5: Power of the VS_{new} test based on small-b asymptotics for the single peak Triangular model for different values of the true visual significance measure. The values in bold correspond to type I error values since the associated VS values are all in the null region (≤ 0.1). The sample size is $n = 600$ and the width δ used in the VS measure is $\pi/15$.

years differs from that for the 16 year average. Thus, we expect to see residual seasonality in last 10 years of data for most of the series after each series has been adjusted for seasonality using 16 year monthly averages, i.e., after replacing each value by its difference from the mean for that month over 16 years. Figure 9 shows the means of first differences of logarithmic data by calendar months for the 16 series where the means are computed based on the last 10 years of data. The plots are all on the same scale. The plots show a varying degree of month to month changes in the mean plots, substantiating the claim of residual seasonality for most of the series.

We applied the VS_{new} test along with the exisiting VS diagnostic procedure in X-12-ARIMA to the last 10 years of data for each of the series. The VS procedure from X-12-ARIMA was applied to the spectrum estimated both using an AR(30) model as well as via the raw periodogram. To compare with other seasonality tests, we consider the Q_s statistics of (X-13-ARIMA 2015; Maravall 2012) and the stable F-test (Lytras, Feldpausch, and Bell (2007)). The Q_s test is a variant of the Box-Ljung-Pierce test applied to seasonal lag autocorrelations. The precise definition of the Q_s statistics can be found in the X-13-ARIMA Reference Manual (X-13-ARIMA 2015; pp-198). The spectral diagnostic tests based on the AR(30) spectrum and periodogram as well as the VS_{new} test are applied to each of the five seasonal frequencies individually. The tests for individual seasonal frequencies were not adjusted for multiple testing. The Q_s statistics and the stable F-tests are aggregate tests over multiple seasonal frequencies.

Table 6.2 reports the peak detection results for all the tests for the 16 series considered. The columns arspec s-pk and pdg s-pk correspond to the X-12-ARIMA VS diagnostics computed based on the AR(30) approximation of the spectrum and the raw periodogram, respectively. The column F^M corresponds to the F-test of Lytras, Feldpausch, and Bell (2007) and the column Q_s corresponds

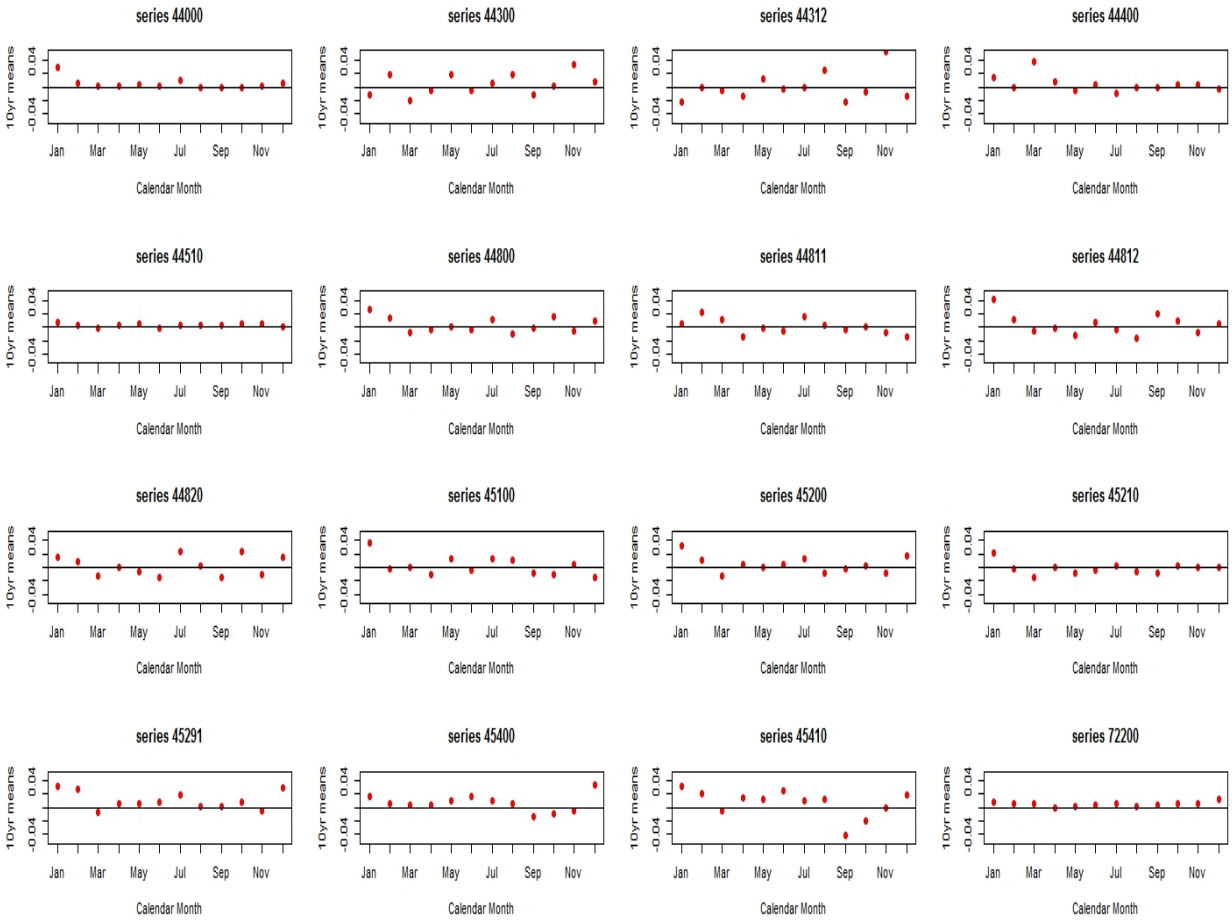


Figure 9: Plots of twelve monthly means, computed over 10 years (red dots) and 16 years (horizontal lines), for 16 retail series.

	Bandwidth fraction b					
VS	.3	.4	.5	.6	.7	.8
0.000	0.001	0.001	0.002	0.002	0.004	0.004
0.050	0.002	0.004	0.005	0.007	0.008	0.007
0.100	0.006	0.009	0.010	0.009	0.010	0.010
0.150	0.015	0.018	0.020	0.021	0.024	0.020
0.200	0.050	0.055	0.057	0.060	0.055	0.050
0.300	0.112	0.121	0.113	0.110	0.113	0.102
0.400	0.270	0.268	0.249	0.234	0.232	0.209
0.500	0.453	0.442	0.419	0.396	0.387	0.349

Table 6: Power of the VS_{new} test based on small-b asymptotics for the single peak Triangular model for different values of the true visual significance measure. The values in bold correspond to type I error values since the associated VS values are all in the null region (≤ 0.1). The sample size is $n = 600$ and the width δ used in the VS measure is $\pi/60$.

	Bandwidth fraction b					
Right VS	.3	.4	.5	.6	.7	.8
0.100	0.011	0.014	0.014	0.015	0.018	0.017
0.300	0.019	0.022	0.024	0.024	0.026	0.022
0.500	0.029	0.036	0.037	0.035	0.039	0.035
0.700	0.038	0.047	0.050	0.050	0.053	0.048
0.900	0.037	0.044	0.047	0.050	0.054	0.052

Table 7: Size of the VS_{new} test based on small-b asymptotics for the asymmetric single peak Triangular model for different values of the true right visual significance measure. The left VS value is fixed at $\tau = 0.1$. The sample size is $n = 600$ and the width δ used in the VS measure is $\pi/60$.

to the test in TRAMO-SEATS. For the aggregates tests (F^M and Q_s) a single p-value is reported for each series. A value of 0.05 indicates that the test was significant at nominal level 0.05 but not at 0.01, while a value of 0.01 indicates that the test was significant at nominal level 0.01. For the spectrum diagnostics and the proposed VS_{new} test the columns list the seasonal frequencies with significant seasonality.

Except for the VS criterion based on the raw periodogram, the diagnostics/tests detect seasonality in most of the series. The total number of series where some seasonality was detected is given at the bottom row. Overall the tests have comparable performance. The VS criterion based on AR spectral estimate detects more peaks than the rest of the tests. The VS_{new} test usually detects a subset of VS peaks as statistically significant peaks with a few discrepancies. The VS_{new} test was insignificant for three series: "Grocery stores","Men's clothing stores" and "Food services and drinking places". A closer look reveals the reason why the VS_{new} test failed to detect seasonality.

	Bandwidth fraction b					
Right VS	.3	.4	.5	.6	.7	.8
0.100	0.053	0.055	0.058	0.055	0.056	0.057
0.300	0.090	0.087	0.090	0.081	0.085	0.084
0.500	0.124	0.119	0.123	0.110	0.112	0.114
0.700	0.143	0.147	0.148	0.133	0.133	0.136
0.900	0.142	0.147	0.151	0.139	0.140	0.141

Table 8: Size of the test based on percentiles of the distribution of minimum of small-b asymptotic pivots for the asymmetric single peak Triangular model for different values of the true right visual significance measure. The left VS value is fixed at $\tau = 0.1$. The sample size is $n = 600$ and the width δ used in the VS measure is $\pi/60$.

	Bandwidth fraction b					
VS	.3	.4	.5	.6	.7	.8
0.100	0.234	0.268	0.287	0.295	0.302	0.310
0.150	0.463	0.500	0.516	0.515	0.513	0.511
0.200	0.705	0.721	0.726	0.722	0.712	0.701
0.300	0.959	0.962	0.961	0.958	0.956	0.952
0.400	0.996	0.996	0.996	0.995	0.995	0.996
0.500	1.000	1.000	1.000	1.000	1.000	1.000

Table 9: Size of test based on the VS rule of rejecting absence of peak hypothesis if the VS functional exceeds $\tau = 0.1$ fraction of the observed dynamic range of log-spectral density. The true VS value is fixed at $\tau = 0.1$. The sample size is $n = 600$ and the width δ used in the VS measure is $\pi/60$.

For the "Grocery stores" series and the "Food services and drinking places" the monthly mean plot in Figure 9 shows that the magnitude of month to month movement in the last 10 years of the adjusted series are minimal, indicating a low degree of moving seasonality in these series. The "Men's clothing stores" series was not a good candidate for seasonal adjustment in the first place, based on the autocorrelation plot and the spectral density estimate. Thus, overall the proposed test has very reasonable performance in detecting residual seasonality.

7 Conclusion

Because seasonal adjustment is an enormous activity for statistical offices, the determination of adjustment inadequacy is extremely important. A host of criteria have been proposed over the decades (summarized in Hylleberg (1986)), and recent work has focused on placing seasonal adjustment diagnostics on a rigorous statistical footing. Following the work of McElroy and Holan (2009), this paper examines the assessment of spectral peaks and incorporates a quantification of

Series	arspec s-pk	pdg s-pk	VS_{new}	F^M	Q_S
44000	1,2,3	1,3	2,4	.01	.01
44300	2,4	–	2,4	.05	.01
44312	3	–	4	–	.01
44400	2,5	–	2	–	.05
44510	1	–	–	–	–
44800	1,3	–	1	.01	.01
44811	2	–	–	–	–
44812	1,3	–	1,2	.05	.01
44820	4	2	1	–	.01
45100	–	–	2	–	.01
45200	1,2,3	1,2	1,2	–	.01
45210	2	1	1	.05	.01
45291	1,2,5	2	1,2,5	–	.01
45400	1,2	1	2	.01	.01
45410	1,2	1	1,2	.01	.01
72200	1	1	–	–	–
Totals	15	8	13	8	13

Table 10: Diagnostic indications of residual seasonality in the last 10 Years of the stable-seasonal adjustment, for 16 retail series delineated in the first column. The second column pertains to the AR(30) VS method of X-12-ARIMA, whereas the third column utilizes the periodogram VS method. The fourth column corresponds to VS_{new}, the methodology of this paper. The fifth column corresponds to the stable F-test, and the sixth column is the Q_s procedure.

Type I error into the decision rule. This advance is achieved by determining the distribution of tapered-spectral estimators and the associated peak functionals, based upon the visual significance criterion of Soukup and Findley (1999).

In order to formulate meaningful peak criteria, our research has explored different parametrizations of spectral peaks, and proposed threshold values for peak functionals to properly capture spectral salients at seasonal frequencies. Given our own choices of peak functionals – largely motivated by the prior literature – we have tabulated the empirical size and power for Bartlett-taper spectral estimators, used a fixed bandwidth fraction asymptotic theory for critical values. A chief finding was that ten years of monthly data requires a fairly broad peak functional in order to obtain the correct size; moreover, for the narrower functional corresponding to the peak functional used in the X-12-ARIMA software, at least twenty years of monthly data is recommended to obtain the correct proportion of Type I errors. Currently the software uses only eight years of data.

Given that most seasonal adjustments arising from the X-12-ARIMA software are adequate, and that moreover incorporating statistical uncertainty into the peak measures makes it *harder* to

detect a peak, we expect in practice that the frequency of incidences of inadequate adjustment will be reduced with the new method. Essentially, the current visual significance procedure has a Type I error rate of zero, and generates false detections of residual seasonality. This expectation has been borne out by our data analysis on four sectors of Census Bureau series. The newer methodology adds greater statistical rigor to the visual significance method, and decreases the incidence of Type I errors.

We make no claim that our proposed peak functionals are in any way optimal, but we have attempted to link them to other notions of seasonality, such as high autocorrelation at seasonal lags. Future research could focus on the specific problem of designing superior peak functionals; if based upon spectral density estimates, the theory of this paper might be applicable. Here we've constrained our investigation to functionals resembling visual significance, in order to allow more direct comparisons with current practice, and develop the theory in a context familiar to many seasonal adjusters. The key takeaway is a caution about apophenia (i.e., the discovery of apparent patterns in chaotic data) in the context of spectral plots: it is easy for the human eye to find peaks that are possibly not sufficiently salient to warrant concern about adjustment adequacy, or – even worse – are actually statistical anomalies due to the estimation process; the latter fallacy is what we seek to ameliorate through the techniques of this paper.

Disclaimer This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Appendix

A.1 Proofs

We now proceed to the proofs of the main theorems, but state two lemmas without proof – the first is proved by application of summation by parts, and the second consists of known trigonometric identities (Gradshteyn and Rhyzik (1994); 1.342). Then we proceed to the proofs of the fixed bandwidth fraction results.

Lemma 1. *Let* $\{\gamma_t\}_{t=1}^n, \{a_t\}_{t=1}^n, \{b_t\}_{t=1}^n$ *be three sequences of scalars. Then*

$$\sum_{t,s=1}^{n} \gamma_{t-s} a_t b_s = \gamma_0 \sum_{t=1}^{n} a_t b_t + \sum_{h=1}^{n-1} \gamma_h \sum_{t=1}^{n-h} [a_t b_{t+h} + b_t a_{t+h}].$$

Lemma 2. *For* $0 \leq \theta \leq 2\pi$,

$$\sum_{t=1}^{n} \cos(\theta t) = \cos((n+1)\theta/2)\, \frac{\sin(n\theta/2)}{\sin(\theta/2)} \tag{A.1}$$

$$\sum_{t=1}^{n} \sin(\theta t) = \sin((n+1)\theta/2)\, \frac{\sin(n\theta/2)}{\sin(\theta/2)}. \tag{A.2}$$

Moreover, suppose that $\theta = \theta_0 + n^{-1}\,[\theta, \theta_0]$ *for some fixed* θ_0. *Then*

$$n^{-1} \sum_{t=1}^{n} \cos(\theta t) = 1_{\{\theta_0 = 0 \ \text{or} \ 2\pi\}} \sin([\theta, \theta_0])/[\theta, \theta_0] + O(n^{-1}), \tag{A.3}$$

$$n^{-1} \sum_{t=1}^{n} \sin(\theta t) = 1_{\{\theta_0 = 0 \ \text{or} \ 2\pi\}} (1 - \cos([\theta, \theta_0]))/[\theta, \theta_0] + O(n^{-1}). \tag{A.4}$$

Proof of Proposition 3. By Lemma 1,

$$\begin{aligned}
2V_{cc}(\theta_j, \theta_k) = {} & \gamma_0 \sum_{t=1}^{n} (\cos([\theta_j + \theta_k]t) + \cos([\theta_j - \theta_k]t)) \\
& + \sum_{h=1}^{n-1} \gamma_h \cos(\theta_j h) \sum_{t=1}^{n-h} (\cos([\theta_j + \theta_k]t) + \cos([\theta_j - \theta_k]t)) \\
& + \sum_{h=1}^{n-1} \gamma_h \cos(\theta_k h) \sum_{t=1}^{n-h} (\cos([\theta_j + \theta_k]t) + \cos([\theta_j - \theta_k]t)) \\
& - \sum_{h=1}^{n-1} \gamma_h \sin(\theta_j h) \sum_{t=1}^{n-h} (\sin([\theta_j + \theta_k]t) + \sin([\theta_j - \theta_k]t)) \\
& - \sum_{h=1}^{n-1} \gamma_h \sin(\theta_k h) \sum_{t=1}^{n-h} (\sin([\theta_j + \theta_k]t) - \sin([\theta_j - \theta_k]t)) \,.
\end{aligned}$$

$$\begin{aligned}
2V_{cs}(\theta_j, \theta_k) = {} & \gamma_0 \sum_{t=1}^{n} (\sin([\theta_j + \theta_k]t) - \sin([\theta_j - \theta_k]t)) \\
& + \sum_{h=1}^{n-1} \gamma_h \cos(\theta_j h) \sum_{t=1}^{n-h} (\sin([\theta_j + \theta_k]t) - \sin([\theta_j - \theta_k]t)) \\
& + \sum_{h=1}^{n-1} \gamma_h \cos(\theta_k h) \sum_{t=1}^{n-h} (\sin([\theta_j + \theta_k]t) - \sin([\theta_j - \theta_k]t)) \\
& + \sum_{h=1}^{n-1} \gamma_h \sin(\theta_j h) \sum_{t=1}^{n-h} (\cos([\theta_j + \theta_k]t) - \cos([\theta_j - \theta_k]t)) \\
& + \sum_{h=1}^{n-1} \gamma_h \sin(\theta_k h) \sum_{t=1}^{n-h} (\cos([\theta_j + \theta_k]t) + \cos([\theta_j - \theta_k]t)) \,.
\end{aligned}$$

$$
\begin{aligned}
2V_{ss}(\theta_j,\theta_k) = \gamma_0 \sum_{t=1}^{n} &(\cos([\theta_j-\theta_k]t) - \cos([\theta_j+\theta_k]t)) \\
&+ \sum_{h=1}^{n-1} \gamma_h \cos(\theta_j h) \sum_{t=1}^{n-h} (\cos([\theta_j-\theta_k]t) - \cos([\theta_j+\theta_k]t)) \\
&+ \sum_{h=1}^{n-1} \gamma_h \cos(\theta_k h) \sum_{t=1}^{n-h} (\cos([\theta_j-\theta_k]t) - \cos([\theta_j+\theta_k]t)) \\
&- \sum_{h=1}^{n-1} \gamma_h \sin(\theta_j h) \sum_{t=1}^{n-h} (\sin([\theta_j-\theta_k]t) - \sin([\theta_j+\theta_k]t)) \\
&+ \sum_{h=1}^{n-1} \gamma_h \sin(\theta_k h) \sum_{t=1}^{n-h} (\sin([\theta_j-\theta_k]t) + \sin([\theta_j+\theta_k]t)) \,.
\end{aligned}
$$

Without loss of generality we derive the limit for the expression $2V_{cc}(\theta_j,\theta_k)$. The other limits will follow analogously. Note that, by Lemma 2,

$$
\begin{aligned}
2n^{-1}V_{cs}(\theta_j,\theta_k) = &\frac{1}{2}\sum_{h=-\infty}^{\infty} \gamma_h(\cos(\theta_j h)+\cos(\theta_k h)) 1_{\{|h|<n\}} \times \\
&[c_{h,n}\big(\frac{\sin(c_{h,n}\,[\theta_j+\theta_k,\theta_j^0+\theta_k^0])}{c_{h,n}\,[\theta_j+\theta_k,\theta_j^0+\theta_k^0]} 1_{\mathcal{A}} + \frac{\sin(c_{h,n}\,[\theta_j-\theta_k,\theta_j^0-\theta_k^0])}{c_{h,n}\,[\theta_j-\theta_k,\theta_j^0-\theta_k^0]} 1_{\mathcal{B}}\big) + O(n^{-1})] \\
&+ \sum_{h=1}^{n-1} \gamma_h \sin(\theta_j h) \sum_{t=1}^{n-h} (\cos([\theta_j+\theta_k]t) - \cos([\theta_j-\theta_k]t)) \\
&+ \sum_{h=1}^{n-1} \gamma_h \sin(\theta_k h) \sum_{t=1}^{n-h} (\cos([\theta_j+\theta_k]t) + \cos([\theta_j-\theta_k]t)) \\
&:= T_{1n} + T_{2n} + T_{3n},
\end{aligned}
$$

where $c_{h,n} = 1 - \frac{|h|}{n}$, and the $O(n^{-1})$ term is uniformly bounded in h and n. Because $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, applying the dominated convergence theorem, we have

$$
\lim_{n\to\infty} T_{1n} = \frac{1}{2}\,[\frac{f(\theta_j^0)+f(\theta_k^0)}{2}] \left(\frac{\sin([\theta_j+\theta_k,\theta_j^0+\theta_k^0])}{[\theta_j+\theta_k,\theta_j^0+\theta_k^0]}\, 1_{\mathcal{A}} + \frac{\sin([\theta_j-\theta_k,\theta_j^0-\theta_k^0])}{[\theta_j-\theta_k,\theta_j^0-\theta_k^0]}\, 1_{\mathcal{B}} \right).
$$

The limits of T_{2n} and T_{3n} can be obtained in similar manner by rearranging the terms in T_{2n} and T_{3n} analogous to T_{1n} and applying the DCT. □

Proof of Theorem 1: The strategy of proof follows that of Theorem 1 of McElroy and Politis (2014). First, $\xi_{[\cdot n]}$ and $S_{[\cdot n]}$ are equivalent processes, for any of the frequencies and for either the sine or cosine transform, in the sense that any linear combination over any set of times of their difference tends to zero in probability (see the proof of Theorem 3 in McElroy and Politis (2013)). So we

can work with the non-linearly interpolated DFTs to show the convergence of finite-dimensional distributions; however, tightness follows from Problem 4.11 of Karatzas and Shreve (1991), since the linearly interpolated processes are in the space of continuous functions – the uniform integrability conditions on the DFTs then yield tightness. For the finite-dimensional distributions, consider the joint convergence of $2J$ components over times $r_1, \cdots, r_m$. Consider

$$G = \sum_{\ell=1}^{m} \alpha_\ell \sum_{j=1}^{J} \left(\eta_j^c S_{[r_\ell n]}^c(\theta_j) + \eta_j^s S_{[r_\ell n]}^s(\theta_j) \right)$$

for arbitrary real numbers α_ℓ and η_j^c, η_j^s. We wish to extend the definitions of V_{cc}, V_{cs}, and V_{ss} by denoting the sample size as superscripts of the two sine/cosine transform arguments to the covariance. That is, write $V_{xy}^{t,s}$ where x, y is either c or s as desired, and t, s denote the number of summands in the x, y transforms respectively. Supposing that $t < s$ for the moment, we have the general identity

$$V_{xy}^{t,s} = \frac{1}{2} \left(V_{xy}^{t,t} + V_{x,y}^{s,s} - V_{xy}^{s-t,s-t} \right), \tag{A.5}$$

which is proved by writing out the double summations involved in the covariance. Now (A.5) expresses the covariances we need in terms of the more standard types studied in Proposition 3. So G has mean zero and variance

$$\begin{aligned} &\sum_{\ell,p=1}^{m} \alpha_\ell \alpha_p \sum_{j,k=1}^{J} (\eta_j^c \eta_k^c V_{cc}^{[r_\ell n],[r_p]n}(\theta_j, \theta_k) + \eta_j^c \eta_k^s V_{cs}^{[r_\ell n],[r_p]n}(\theta_j, \theta_k) \\ &+ \eta_j^s \eta_k^c V_{sc}^{[r_\ell n],[r_p]n}(\theta_j, \theta_k) + \eta_j^s \eta_k^s V_{ss}^{[r_\ell n],[r_p]n}(\theta_j, \theta_k)), \end{aligned}$$

and we can apply (A.5) to the individual summands. Supposing that $r_\ell < r_p$ for the moment, we have

$$\begin{aligned} V_{xy}^{[r_\ell n],[r_p n]} &= \frac{1}{2} \left(V_{xy}^{[r_\ell n],[r_\ell n]} + V_{x,y}^{[r_p n],[r_p n]} - V_{xy}^{[r_p n]-[r_\ell n],s-[[r_p n]_\ell n]} \right) \\ &= o(n) + \frac{1}{2} \left(r_\ell n \underline{V}_{xy} + r_p n \underline{V}_{xy} - (r_p - r_\ell) n \underline{V}_{xy} \right) \\ &= o(n) + r_\ell n \underline{V}_{xy}. \end{aligned}$$

This uses Proposition 3. So applying this to the variance of G, we obtain up to $o(n)$ terms

$$n \sum_{\ell,p=1}^{m} \alpha_\ell \alpha_p (r_\ell \wedge r_p) \sum_{j,k=1}^{J} (\eta_j^c \eta_k^c \underline{V}_{cc}(\theta_j, \theta_k) + \eta_j^c \eta_k^s \underline{V}_{cs}(\theta_j, \theta_k) + \eta_j^s \eta_k^c \underline{V}_{sc}(\theta_j, \theta_k) + \eta_j^s \eta_k^s \underline{V}_{ss}).$$

Recalling the definition of the vector $A(r)$ process from the theorem's statement, and with η composed of the various η_j^c, η_j^s, and using the properties of Brownian Motion, the above expression is seen to be equal to (up to terms $o(n)$) n times the variance of $\sum_{\ell=1}^{m} \alpha_\ell \eta' A(r_\ell)$, which verifies the convergence of variances. If P1 holds, the asymptotic normality of the standardized partial sums follows from Theorem 5.2.3 of Taniguchi and Kakizawa (2000), given the above convergence

of variances. Under P2 utilize Theorem 5.2.2 of Taniguchi and Kakizawa (2000) for short memory, using the convergence of variances. Under P3 the conditions on higher order cumulants can be utilized to show that higher order cumulants of the DFT will grow less rapidly then the normalizing rate, and hence are asymptotically negligible. This completes the proof. □

Proof of Theorem 2: This follows immediately from Theorem 1 once we utilize the equivalency between the linearly interpolated sine and cosine transforms and the original DFTs, along the lines of the proof of Theorem 2 of McEroy and Politis (2014). In detail, the tapered acf spectral estimatorr at a frequency θ_j can be written in terms of the sine and cosine transformation processes at that frequency as

$$\begin{aligned} n\widehat{f}_b(\theta_j) &= 2S_n^c(\theta_j)\int_0^1 S^c_{\lfloor rn\rfloor}(\theta_j)\partial_n\Lambda_b(1-r)dr + \Lambda_b(0)S_n^c(\theta_j)^2 \\ &\quad -\int_0^1\int_0^1 S^c_{\lfloor rn\rfloor}(\theta_j)\partial_n^2\Lambda_b(r-s)S^c_{\lfloor rn\rfloor}(\theta_j)drds \\ &\quad +2S_n^s(\theta_j)\int_0^1 S^s_{\lfloor rn\rfloor}(\theta_j)\partial_n\Lambda_b(1-r)dr + \Lambda_b(0)S_n^s(\theta_j)^2 \\ &\quad -\int_0^1\int_0^1 S^c_{\lfloor rn\rfloor}(\theta_j)\partial_n^2\Lambda_b(r-s)S^c_{\lfloor rn\rfloor}(\theta_j)drds \end{aligned} \tag{A.6}$$

where

$$\begin{aligned} \partial_n\Lambda_b(r) &= n\Big[\Lambda_b(\frac{\lfloor rn\rfloor+1}{n}) - \Lambda_b(\frac{\lfloor rn\rfloor}{n})\Big], \\ \partial_n^2\Lambda_b(r) &= n^2\Big[\Lambda_b(\frac{\lfloor rn\rfloor+1}{n}) - 2\Lambda_b(\frac{\lfloor rn\rfloor}{n}) + \Lambda_b(\frac{\lfloor rn\rfloor-1}{n})\Big]. \end{aligned}$$

Then we apply the functional limit theorem, carefully attending to any jump discontinuities in the taper, as discussed in the expanded version (technical report online) of McElroy and Politis (2014). □

References

[1] Bell, W.R. and Hillmer, S. C. (1983). Modeling time series with calendar variation. *Journal of the American Statistical assoc.,* **78**, 526–534.

[2] Berger, R. L. (1989). Uniformly more powerful tests for hypotheses concerning linear inequalities and normal means. *Journal of the American Statistical Association,* **84**, 192–199.

[3] Berk, K. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* **2**, 489–502.

[4] Dahlhaus, R. (1983). Spectral analysis with tapered data. *J. Time Ser. Anal.* **4**, 163–175.

[5] Dahlhaus, R. (1985). Asymptotic normality of spectral estimates. *J. Multivariate Anal.,* **16**, 412–431.

[6] Fase, M.M., Konig, J., and Volgenants, A.F. (1973). An experimental look at seasonal adjustment. *De Economist* **121**, 441–486.

[7] Findley, D.F. (2005) Some recent developments and directions in seasonal adjustment. *Journal of Official Statistics* **21**, 343–365.

[8] Gradshteyn, I.S. and Ryzhik, I.M. (1994). *Table of Integrals, Series, and Products, seventh edition*. New York: Academic Press.

[9] Grether, D. M., and Nerlove, M. (1970). Some Properties of 'Optimal' Seasonal Adjustment. *Econometrica,* **38** 682–703.

[10] Hashimzade, N. and Vogelsang, T. (2008). Fixed-b asymptotic approximation of the sampling behaviour of nonparametric spectral density estimators. *Journal of Time Series Analysis,* **29**, 142–162.

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

[12] Hylleberg, S. (1986). *Seasonality in Regression*. Orlando, Florida: Academic Press.

[13] Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus*. Springer, New York.

[14] Lytras, D. P., Feldpausch, R. M., and Bell, W. R. (2007). Determining Seasonality: A Comparison of Diagnostics From X-12-ARIMA. *Proceedings of the Third International Conference on Establishment Surveys.* http://www.census.gov/srd/www/sapaper/sapaper.html

[15] Liu, H. and Berger, R. L. (1995). Uniformly More Powerful, One-Sided Tests for Hypotheses About Linear Inequalities. *Annals of Statistics,* **23**, 55–72.

[16] Maravall, A. (2012). Update of Seasonality Tests and Automatic Model Identification in TRAMO-SEATS. Bank of Spain.

[17] McElroy, T. and Holan, S. (2009). A Nonparametric Test for Residual Seasonality. *Survey Methodology* **35**, 67 – 83.

[18] McElroy, T. and Politis, D. (2013) Distribution Theory for the Studentized Mean for Long, Short, and Negative Memory Time Series. *Journal of Econometrics* **177**, 60–74.

[19] McElroy, T. and Politis, D. (2014). Spectral Density and Spectral Distribution Inference for Long Memory Time Series via Fixed-b Asymptotics. *J. of Econometrics,* **182,** 211–225.

[20] Nerlove, Marc. (1964). Spectral Analysis of Seasonal Adjustment Procedures. *Econometrica,* **32** 241–286.

[21] Parzen, E. (1957). On Consistent Estimates of the Spectrum of a Stationary Time Series. *Annals of Math. Stat.,* **28**, 329–348.

[22] Pierce, D.A. (1976). Uncertainty in Seasonal Adjustment Procedures. *Proceedings of the Business and Economic Statistics Section of the American Statistical Association.* Washington, D.C. 528533.

[23] Pierce, D.A. (1979). Seasonal Adjustment When Both Deterministic and Stochastic Seasonality are Present. *Seasonal Analysis of Economic Time Series, Ed. Arnold Zellner,* p. 242 - 280 NBER.

[24] Priestley, M. B. (1981). *Spectral Analysis and Time Series.* New York: Academic Press.

[25] Sasabuchi, S. (1980). A test of a multivariate normal mean with composite hypotheses determined by linear inequalities. *Biometrika,* **67**, 429–439.

[26] Seasonal Adjustment Centre of Competence (2015). http://www.cros-portal.eu/content/seasonal-adjustment

[27] Soukup, R. and Findley, D. (1999). On the spectrum diagnostics used by X-12-ARIMA to indicate the presence of trading day effects after modeling or adjustment. In *Proceedings of the Third International Conferene on Establishment Surveys (ICES-III)*, Montreal, Canada.

[28] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag, New York City, New York.

[29] Tiao, G. and Tsay, R. (1983). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Annals of Statistics* **11**, 856–871.

[30] X-13-ARIMA (2015). X-13-ARIMA Reference Manual, U.S. Census Bureau, Washington D.C. USA, (Available from https://www.census.gov/ts/x13as/docX13AS.pdf.)