Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

A nonparametric test for residual seasonality

by Tucker McElroy and Scott Holan

June 2009

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2009
Vol. 35, No. 1, pp. 67-83
Statistics Canada, Catalogue No. 12-001-X

A nonparametric test for residual seasonality

Tucker McElroy and Scott Holan [1]

Abstract

Peaks in the spectrum of a stationary process are indicative of the presence of stochastic periodic phenomena, such as a stochastic seasonal effect. This work proposes to measure and test for the presence of such spectral peaks via assessing their aggregate slope and convexity. Our method is developed nonparametrically, and thus may be useful during a preliminary analysis of a series. The technique is also useful for detecting the presence of residual seasonality in seasonally adjusted data. The diagnostic is investigated through simulation and an extensive case study using data from the U.S. Census Bureau and the Organization for Economic Co-operation and Development (OECD).

Key Words: Multiple testing; Nonparametric density estimation; Seasonal adjustment; Spectral density.

1. Introduction

The presence of a peak in the spectrum of a stationary process is indicative of periodic behavior, such as seasonality or a trading day effect. There is a widespread interest in the identification of such peaks in the engineering and econometrics literature, since a pronounced spectral node will exert a potent influence on the dynamics of the stochastic process. A peak indicates a range of frequencies that offer a relatively large contribution to the overall variance of the stochastic process. If the strength of the peak, assessed through its height and width relative to neighboring values, is sufficiently pronounced, any model of the dynamics that ignores the corresponding periodicities will be misspecified. In both engineering and econometrics, one may be interested in signal extraction or forecasting, both of which are sensitive to the presence of spectral peaks.

By a spectral peak, we refer to a region of the spectral density that has greater spectral mass than its immediate neighbors; a more precise definition is developed below. Due to the applications that we have in mind, our peaks have finite height, and thus correspond to stochastic periodic effects in a stationary process. Thus, we are not principally concerned with the detection of fixed (deterministic) periodic effects, nor with nonstationary periodic phenomena (though we make some extensions to this case in Section 3.4 below), as both of these correspond to a spectral peak with infinite height. The vast literature dealing with the detection of fixed effects is discussed in Priestley (1981); for our applications the periodic aspects of the data are not fixed, but instead evolve over time.

In this paper we focus on the application to seasonal adjustment. Specifically, we concentrate on so-called seasonal peaks, which may occur at the seasonal frequencies (assuming a monthly sampling interval) $\pi/6,\ 2\pi/6,\ 3\pi/6,\ 4\pi/6,\ 5\pi/6,$ and $6\pi/6$. The detection of seasonality and residual seasonality presents an important practical problem in federal statistics, and the spectrum is a natural tool towards this end. The frequency domain approach to the detection and analysis of seasonality enjoys wide popularity, because it provides a very natural way to view quasi-periodic behavior. In fact, seasonality is − informally speaking − characterized by the presence of at least one seasonal peak in the spectrum (Nerlove 1964). Frequency domain methods are now employed in X-12-ARIMA (Findley, Monsell, Bell, Otto and Chen 1998) and are part of TRAMO-SEATS (Maravall and Caporello 2004), the two most widely-used seasonal adjustment programs available to the public. Note that frequency domain methods can be implemented via either a parametric (*i.e.*, model-based) or nonparametric approach. We develop a non-parametric diagnostic, which can be invoked to determine the efficacy of *any* seasonal adjustment procedure, either model-based or nonparametric. As noted in Findley, Monsell, Bell, Otto, and Chen (1998), the use of fixed periodic functions alone to model seasonality is typically inadequate for economic data (also see the discussion in Bell and Hillmer 1984).

Spectral peaks at seasonal frequencies in a seasonally adjusted series may indicate inadequacy of the seasonal filters − see Soukup and Findley (1999) for a discussion. At a minimum, seasonal adjustment filters should remove *nonstationary* seasonality and any fixed periodic effects − those phenomena in the observed series that contribute a seasonal pole to the spectrum. However, there is a consensus among seasonal adjusters that it is also desirable to remove some aspects of the *stationary* seasonality as well − hence the explosion of effort in developing model-based seasonal adjustment filters (Bell and Hillmer 1984).

1. Tucker McElroy, Statistical Research Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100. E-mail: tucker.s.mcelroy@census.gov; Scott Holan, Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO, 65211-6100. E-mail: holans@missouri.edu.

The most important prior literature on this topic is Soukup and Findley (1999), which proposes using an autoregressive spectrum to find "visually significant" peaks – essentially the value of the spectrum at each seasonal frequency (or trading day frequency) is compared to its nearest neighbors, and is classified as a peak if the discrepancy is suitably large. This method is currently implemented in the X-12-ARIMA program from the U.S. Census Bureau (2002). One limitation of this approach is that it has really no statistical component: the significance is not statistical – *i.e.*, it is not associated with a hypothesis test – and the thresholds to determine "visual significance" are determined in an *ad hoc* fashion. This paper provides a statistical significance test for peak detection, and can thus be used to offer supplementary statistical evidence of the presence of a peak.

Another related paper is Newton and Pagano (1983), which develops consistent estimators for the local maximizers of the spectrum. Our approach is slightly different, in that we already know the frequencies of interest (the six seasonal frequencies) but seek to test for the presence of a statistically significant peak. Viewing the true spectral density f as a smooth function (this can be quantified through sufficiently rapid decay of the autocovariance function), a peak is a frequency λ_0 such that

$$\dot{f}(\lambda_0) = 0 \qquad \ddot{f}(\lambda_0) < 0, \tag{1}$$

where $\dot{f}$ and $\ddot{f}$ denote first and second derivatives. Clearly, the second derivative must be negative *with some significance* in order for the concept to be meaningful. Upon further reflection, it seems that examining the infinitesimal geometry of f at the single point λ_0 is naïve, since any small spike in the side of a monotonic function may satisfy (1) while being dissociated from more intuitive notions of what constitutes a peak. Therefore, we must have negative convexity in a reasonably large neighborhood of λ_0. This thinking leads to the diagnostic of this paper: aggregate measures of the slope and convexity of the spectral density, appropriately normalized. Mathematically, these will take the form of kernel-smoothed periodogram estimates, but without the bandwidth being dependent on sample size.

In Section 2 we develop the mathematical ideas of this method, illustrated through two carefully chosen choices of kernels. Section 3 shows how statistical estimators can be formulated, and how statistical peak hypotheses can be tested. The methodology is tested in Section 4; simulations provide a finite sample description of the size and power of our test. We further demonstrate the utility of our methods through an extensive case study involving 130 time series from the U.S. Census Bureau and the Organization for Economic Co-operation and Development (OECD). We use some concepts from the multiple testing literature (Hochberg 1988) to combine tests based on the individual frequencies together into one diagnostic. Section 5 concludes, and all theorems and proofs are left to the Appendix.

2. Measuring the local geometry of the spectrum

We begin by discussing the geometry of the spectral density (or spectrum) of the time series under consideration. The starting point is to consider measures of slope and convexity of the spectrum that are completely deterministic (*cf.* the approach of Newton and Pagano 1983); later in Section 3 we will consider statistical measures. In Section 2.1 we introduce the concepts of slope and convexity measures. The relevancy of these measures to peak identification is discussed in 2.2, while 2.3 provides two simple kernels as explicit examples.

Suppose that, after suitable transformations and differencing if necessary, $X_1, X_2, ..., X_n$ is a sample from a zero-mean stationary stochastic process. We will use the notation $X = (X_1, X_2, ..., X_n)'$. The spectral density $f(\lambda)$ is well-defined so long as the autocovariance function $\gamma_f(h)$ is absolutely summable, and is given by

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_f(h) e^{-ih\lambda} \tag{2}$$

with $i = \sqrt{-1}$ and $\lambda \in [-\pi, \pi]$. It follows that the inverse Fourier transform yields

$$\gamma_f(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) e^{ih\lambda} d\lambda, \tag{3}$$

a relation that we will use repeatedly in the sequel. Of course this relationship between γ_g and g holds for any integrable function g, not just a spectral density. Furthermore, denoting the Toeplitz matrix associated with γ_g by $\Sigma(g)$, it follows that

$$\Sigma_{jk}(g) = \gamma_g(j - k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) e^{i(j-k)\lambda} d\lambda.$$

Now from (2), f is d times continuously differentiable if $\sum_{h=-\infty}^{\infty} |h|^d |\gamma_f(h)| < \infty$. We assume that f is twice continuously differentiable for the remainder of the paper (this space of functions will be abbreviated as C^2).

2.1 Measures of slope and convexity

The local geometry of a C^2 function can be described through its first and second derivatives; an aggregate measure of these derivatives is obtained by integrating over a band of frequencies. Alternatively, one may integrate against a function A that has compact support over this band, so long as A provides a suitable proxy for integration over the band. We denote this integral via the general device of a functional θ_A, where

$$\theta_A(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda) f(\lambda) d\lambda. \quad (4)$$

The function A will be referred to as the "kernel" of this functional. Hence the aggregate slope and convexity measures are defined by $\theta_A(\dot{f})$ and $\theta_A(\ddot{f})$, where each dot denotes a single derivative. These functionals give a summary measure of slope and convexity of f over some band $[\mu - \beta/2, \mu + \beta/2] \subset [0, \pi]$, and the corresponding kernels will therefore be denoted $A_{\beta,\mu}$. We consider kernels with the following properties: (i) $A_{\beta,\mu}$ is a C^2 function on $[-\pi, \pi]$; (ii) $A_{\beta,\mu}$ is zero outside the band $[\mu - \beta/2, \mu + \beta/2]$; (iii) $A_{\beta,\mu}$ is symmetric about μ on this band; (iv) $\dot{A}_{\beta,\mu}(\mu \pm \beta/2) = 0$. Condition (iii) ensures that the location of the peak in f is not shifted by employing the kernel $A_{\beta,\mu}$. Note that we do not impose that the total integral of $A_{\beta,\mu}$ be unity, because later we will employ a normalization that will automatically account for the total mass of the kernel. Now by (iv) and integration by parts in (4), we obtain

$$\theta_{A_{\beta,\mu}}(\dot{f}) = -\theta_{\dot{A}_{\beta,\mu}}(f)$$
$$\theta_{A_{\beta,\mu}}(\ddot{f}) = \theta_{\ddot{A}_{\beta,\mu}}(f). \quad (5)$$

These formulas are convenient, because they only require a knowledge of f, not its derivatives (assuming that we can compute $\dot{A}_{\beta,\mu}$ and $\ddot{A}_{\beta,\mu}$). Following the extensive literature on kernels in nonparametric regression and spectral density estimation, we can start with an even kernel A defined on the band $[-\pi, \pi]$ that satisfies (i) and $\dot{A}(\pm\pi) = 0$. Then $A_{\beta,\mu}$ is defined via

$$A_{\beta,\mu}(\lambda) = \frac{2\pi}{\beta} A\left(\frac{2\pi}{\beta}(\lambda - \mu)\right),$$

and is zero outside the band of frequencies $[\mu - \beta/2, \mu + \beta/2]$. Clearly we must impose $\beta \le 2\mu$ and $\beta \le 2(\pi - \mu)$, so that $[\mu - \beta/2, \mu + \beta/2] \subset [0, \pi]$; and the kernel $A_{\beta,\mu}$ satisfies conditions (i)-(iv). Note that we cannot construct these types of measures for μ equal to 0 or π. Using a change of variables, we see that

$$\gamma_{A_{\beta,\mu}}(h) = \exp\{ih\mu\}\gamma_A(h\beta/2\pi), \quad (6)$$

so that the effect of β and μ are in some sense separable. Note that we typically evaluate γ_A at non-integer values, so these relations are obtained by extending (3) to non-integer arguments. The fact that $\gamma_{A_{\beta,\mu}}$ is complex-valued may seem troubling, but actually only its real portion will enter into our statistical estimators. Of course, we are ultimately interested in $\dot{A}_{\beta,\mu}$ and $\ddot{A}_{\beta,\mu}$, which are given by

$$\dot{A}_{\beta,\mu}(\lambda) = \frac{4\pi^2}{\beta^2}\dot{A}\left(\frac{2\pi}{\beta}(\lambda - \mu)\right),$$

and

$$\ddot{A}_{\beta,\mu}(\lambda) = \frac{8\pi^3}{\beta^3}\ddot{A}\left(\frac{2\pi}{\beta}(\lambda - \mu)\right).$$

Later, we will consider the squares of such kernels, and their corresponding inverse Fourier transforms. Hence assuming that $[\mu - \beta/2, \mu + \beta/2] \subset [0, \pi]$, the squares are given by

$$\dot{A}^2_{\beta,\mu}(\lambda) = \frac{16\pi^4}{\beta^4}\dot{A}^2\left(\frac{2\pi}{\beta}(\lambda - \mu)\right)$$

and

$$\ddot{A}^2_{\beta,\mu}(\lambda) = \frac{64\pi^6}{\beta^6}\ddot{A}^2\left(\frac{2\pi}{\beta}(\lambda - \mu)\right).$$

Finally, we notice from (4) that we can rewrite $\theta_A(f)$ as

$$\theta_A(f) = \sum_{h=-\infty}^{\infty} \gamma_A(h)\gamma_f(h). \quad (7)$$

Thus it may be advantageous to determine the $\gamma_A(h)$ sequence from the kernel A. Taking the inverse Fourier Transform of the above slope and convexity kernels, we can construct $\Sigma(\dot{A}_{\beta,\mu})$, $\Sigma(\ddot{A}_{\beta,\mu})$, $\Sigma(\dot{A}^2_{\beta,\mu})$ and $\Sigma(\ddot{A}^2_{\beta,\mu})$, as follows:

$$\gamma_{\dot{A}_{\beta,\mu}}(h) = \frac{2\pi}{\beta}\exp\{ih\mu\}\gamma_{\dot{A}}(h\beta/2\pi),$$
$$\gamma_{\ddot{A}_{\beta,\mu}}(h) = \frac{4\pi^2}{\beta^2}\exp\{ih\mu\}\gamma_{\ddot{A}}(h\beta/2\pi),$$
$$\gamma_{\dot{A}^2_{\beta,\mu}}(h) = \frac{8\pi^3}{\beta^3}\exp\{ih\mu\}\gamma_{\dot{A}^2}(h\beta/2\pi),$$
$$\gamma_{\ddot{A}^2_{\beta,\mu}}(h) = \frac{32\pi^5}{\beta^5}\exp\{ih\mu\}\gamma_{\ddot{A}^2}(h\beta/2\pi). \quad (8)$$

Thus, if we have the time-domain information $\gamma_f(h)$ for the process $\{X_t\}$, we can compute slope and convexity measures using (7) given the inverse Fourier transform sequence of the appropriate kernels. Since $\gamma_f(h)$ is a symmetric sequence, we only need to consider the real portion of $\gamma_A(h)$ if it happens to be complex.

2.2 Troughs and peaks

The aggregate measures of spectral slope and convexity previously described provide the building blocks for determinants of the local spectral geometry. Our overall interest is in determining whether a given interval of the spectrum is a peak or a trough (or is monotonic). In the second order geometry of calculus, a local maximum has the defining property that the first derivative is zero and the

second derivative is strictly negative. Obviously this requires looking, sequentially, at a slope measure and a convexity measure, defined over the same band of frequencies.

In order to test for the presence of a peak, the sequential approach can be seen as making inferential statements about $\theta_{A_{\beta,\mu}}(\dot{f})$ and $\theta_{A_{\beta,\mu}}(\ddot{f})$. Note, in making these inferential statements we choose μ ahead of time, according to where in the spectrum we wish to detect a peak (or trough); β is chosen according to which frequencies we wish to exclude, a decision based on how local we wish our viewpoint of the spectrum to be. Then we say that μ is a β-aggregate peak (with respect to A) of the spectrum if

$$\theta_A(\dot{f}) = 0 \quad \text{and} \quad \theta_A(\ddot{f}) < 0.$$

The sequential aspect comes from the idea that we generally determine whether $\theta_A(\dot{f}) = 0$ first, and then determine the convexity; this will become more apparent when we consider statistical testing in Section 3.2. In a similar manner we define a β-aggregate trough when $\theta_A(\ddot{f}) > 0$. In terms of hypothesis testing for a peak, we have

$$H_0^{(1)}\colon \theta_A(\dot{f}) = 0 \quad vs. \quad H_a^{(1)}\colon \theta_A(\dot{f}) \neq 0$$
$$H_0^{(2)}\colon \theta_A(\ddot{f}) = 0 \quad vs. \quad H_a^{(2)}\colon \theta_A(\ddot{f}) < 0.$$

The unusual aspect of this hypothesis test is that we wish to fail to reject $H_0^{(1)}$ first, and then conditional on this test we want to reject $H_0^{(2)}$ in favor of the alternative $H_a^{(2)}$.

2.3 Examples of kernels

There are a host of kernels that satisfy conditions (i) through (iv); we can simply borrow from the literature on nonparametric density estimation. For example, the Parzen and Tukey-Hanning (TH) kernels (discussed in Priestley 1981) are suitable, whereas the Bartlett and Daniell kernels are inappropriate, since (iv) does not hold. In general, one only needs to use (8) to determine the inverse Fourier transforms. In this section, we consider two examples: Quartic and TH. The advantage of these kernels is that they have easily computable first and second derivatives, and their inverse Fourier transforms can be obtained explicitly.

Example 1: *Quartic Kernel*

We begin by considering a polynomial kernel of degree four, namely a quartic. Imposing all of the constraints (i) through (iv) yields the following form:

$$A(\lambda) = \frac{15}{8\pi^4}(\lambda^4 - 2\pi^2\lambda^2 + \pi^4),$$
$$\dot{A}(\lambda) = \frac{15}{8\pi^4}(4\pi^3 - 4\pi^2\lambda), \text{ and}$$
$$\ddot{A}(\lambda) = \frac{15}{8\pi^4}(12\pi^2 - 4\pi^2).$$

Taking the inverse Fourier transform of the slope and convexity kernels (and their squares) yields

$$\gamma_{\dot{A}}(h) = \frac{15i}{\pi^5}\left(\frac{\pi^2\sin\pi h}{h^2} + \frac{3\pi\cos\pi h}{h^3} - \frac{3\sin\pi h}{h^4}\right),$$
$$\gamma_{\ddot{A}}(h) = \frac{15}{\pi^5}\left(\frac{\pi^2\sin\pi h}{h} + \frac{3\pi\cos\pi h}{h^2} - \frac{3\sin\pi h}{h^3}\right),$$
$$\gamma_{\dot{A}^2}(h) = \frac{225}{\pi^9}\left(-\frac{2\pi^4\sin\pi h}{h^3} - \frac{18\pi^3\cos\pi h}{h^4} + \frac{78\pi^2\sin\pi h}{h^5} + \frac{180\pi\cos\pi h}{h^6} - \frac{18\sin\pi h}{h^7}\right),$$
$$\gamma_{\ddot{A}^2}(h) = \frac{225}{4\pi^{11}}\left(\frac{\pi^4\sin\pi h}{h} + \frac{6\pi^3\cos\pi h}{h^2} - \frac{24\pi^2\sin\pi h}{h^3} - \frac{54\pi\cos\pi h}{h^4} + \frac{54\sin\pi h}{h^5}\right),$$

to which we apply (8) and obtain

$$\gamma_{\dot{A}_{\beta,\mu}}(h) = \frac{30i}{\beta}\exp\{ih\mu\}\left(\frac{\sin k}{k^2} + \frac{3\cos k}{k^3} - \frac{3\sin k}{k^4}\right),$$
$$\gamma_{\ddot{A}_{\beta,\mu}}(h) = \frac{30}{\beta^2\pi}\exp\{ih\mu\}\left(\frac{\sin k}{k} + \frac{3\cos k}{k^2} - \frac{3\sin k}{k^3}\right),$$
$$\gamma_{\dot{A}^2_{\beta,\mu}}(h) = \frac{1{,}800\pi}{\beta^3}\exp\{ih\mu\}\left(\frac{2\sin k}{k^3} + \frac{18\cos k}{k^4} - \frac{78\sin k}{k^5} - \frac{180\cos k}{k^6} + \frac{180\sin k}{k^7}\right), \text{ and}$$
$$\gamma_{\ddot{A}^2_{\beta,\mu}}(h) = \frac{1{,}800}{\beta^5\pi}\exp\{ih\mu\}\left(\frac{\sin k}{k} + \frac{6\cos k}{k^2} - \frac{24\sin k}{k^3} - \frac{54\cos k}{k^4} + \frac{54\sin k}{k^5}\right),$$

where $k = h\beta/2$. Note that $\gamma_{\dot{A}^2_{\beta,\mu}}(0) = 240\pi/(7\beta^3)$ and $\gamma_{\ddot{A}^2_{\beta,\mu}}(0) = 360/(\beta^5\pi)$ follow by application of L'Hopital's rule. These formulas allow us to construct the appropriate Toeplitz matrices for the diagnostic (as discussed in Section 3.1 below, it suffices to consider the real part of these sequences).

Example 2: *TH Kernel*

A similar shape to the quartic can be obtained through the use of a cosine function. The following choice satisfies all the stated conditions on a kernel:

$$A(\lambda) = \frac{1}{2\pi}(1 + \cos\lambda),$$
$$\dot{A}(\lambda) = \frac{1}{2\pi}(-\sin\lambda), \text{ and}$$
$$\ddot{A}(\lambda) = \frac{1}{2\pi}(-\cos\lambda).$$

This function is identical to the Tukey-Hanning lag window, though here we apply it as a spectral window (see Priestley 1981). Hereafter it will be referred to as the TH kernel. Taking the inverse Fourier transform of the slope and convexity kernels (and their squares) yields

$$\gamma_{\dot{A}}(h)=\frac{i}{4\pi^2}\left(\frac{\sin\pi(h+1)}{h+1}-\frac{\sin\pi(h-1)}{h-1}\right),$$

$$\gamma_{\ddot{A}}(h)=-\frac{1}{4\pi^2}\left(\frac{\sin\pi(h+1)}{h+1}+\frac{\sin\pi(h-1)}{h-1}\right),$$

$$\gamma_{\dot{A}^2}(h)=\frac{1}{16\pi^3}\left(\frac{2\sin\pi h}{h}-\frac{\sin\pi(h+2)}{h+2}-\frac{\sin\pi(h-2)}{h-2}\right), \text{ and}$$

$$\gamma_{\ddot{A}^2}(h)=\frac{1}{16\pi^3}\left(\frac{2\sin\pi h}{h}+\frac{\sin\pi(h+2)}{h+2}+\frac{\sin\pi(h-2)}{h-2}\right).$$

Now applying (8) yields

$$\gamma_{\dot{A}_{\beta,\mu}}(h) = \frac{i}{2\beta}\exp\{ih\mu\}\left(\frac{\sin(k+\pi)}{k+\pi}-\frac{\sin(k-\pi)}{k-\pi}\right),$$

$$\gamma_{\ddot{A}_{\beta,\mu}}(h) = -\frac{\pi}{\beta^2}\exp\{ih\mu\}\left(\frac{\sin(k+\pi)}{k+\pi}+\frac{\sin(k-\pi)}{k-\pi}\right),$$

$$\gamma_{\dot{A}^2_{\beta,\mu}}(h) = \frac{\pi}{2\beta^3}\exp\{ih\mu\}\left(\frac{2\sin k}{k}-\frac{\sin(k+2\pi)}{k+2\pi}-\frac{\sin(k-2\pi)}{k-2\pi}\right), \text{ and}$$

$$\gamma_{\ddot{A}^2_{\beta,\mu}}(h) = \frac{2\pi^3}{\beta^5}\exp\{ih\mu\}\left(\frac{2\sin k}{k}-\frac{\sin(k+2\pi)}{k+2\pi}+\frac{\sin(k-2\pi)}{k-2\pi}\right),$$

where $k = h\beta/2$. Note that $\gamma_{\dot{A}^2_{\beta,\mu}}(0) = \pi/\beta^3$ and $\gamma_{\ddot{A}^2_{\beta,\mu}}(0) = 4\pi^3/\beta^5$ follow by application of L'Hopital's rule (using the convention that $\sin(0)/0 = 1$). These formulas allow us to construct the appropriate Toeplitz matrices for the diagnostic (again, as discussed in Section 3.1 below, it suffices to consider the real part of these sequences).

3. Statistical methodology

Of course we do not typically have knowledge of the spectrum f, and thus it is usually necessary to form estimates from the data. In this section we describe statistical estimates of slope and convexity measures that are consistent and simple to compute in the time-domain. Under some mild additional assumptions, these estimates are asymptotically normal, which will be advantageous when performing hypothesis tests. In Section 3.1 the statistical estimates are defined, and their asymptotic properties are discussed. Section 3.2 discusses the application to peak testing, and 3.3 gives an extension to joint peak testing, which facilitates an important application in seasonal adjustment. Section 3.4 discusses extensions to trend nonstationary data.

3.1 Estimators of slope and convexity

We begin by noting that the quadratic form (for any integrable function g)

$$\frac{1}{n}X'\Sigma(g)X = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda)\, I(\lambda)\, d\lambda,$$

where I denotes the periodogram. Although the periodogram is typically defined at the Fourier frequencies $(2\pi j/n;\ j = 1, ..., \lfloor n/2 \rfloor)$ we define it at a continuous band of frequencies as follows

$$I(\lambda) = \frac{1}{n}\left|\sum_{t=1}^{n} X_t e^{-it\lambda}\right|^2 = \sum_{h=1-n}^{n-1} R(h)\, e^{-it\lambda},\ \lambda \in [-\pi, \pi] \quad (9)$$

with $R(h)$ equal to the sample (uncentered) autocovariance function. This gives an elegant way of passing from the time-domain to the frequency-domain, and is well-known in the time series literature (see Taniguchi and Kakizawa 2000). Moreover, such integrals of the periodogram are generally consistent, *i.e.*, $\theta_g(I) \xrightarrow{a.s.} \theta_g(f)$ as $n \to \infty$, under mild conditions discussed below (note that the inconsistency of the periodogram is resolved by the spectral aggregation against the function g, as shown in the Appendix). Therefore, we obtain statistical estimates of the slope and convexity measures f by using a "plug in" approach, *i.e.*, we simply replace f by I in $\theta_{A_{\beta,\mu}}$. In particular,

$$\hat{\theta}_{A_{\beta,\mu}}(\dot{f}) = -\theta_{\dot{A}_{\beta,\mu}}(I) = -\frac{1}{n}X'\Sigma(\dot{A}_{\beta,\mu})X, \text{ and}$$

$$\hat{\theta}_{A_{\beta,\mu}}(\ddot{f}) = \theta_{\ddot{A}_{\beta,\mu}}(I) = \frac{1}{n}X'\Sigma(\ddot{A}_{\beta,\mu})X. \quad (10)$$

This definition makes use of (5), which accounts for the minus sign in the slope measure. In order to compute the estimate, we utilize the time-domain representation (expressed as a quadratic form). This representation is convenient in that we only need determine a suitable length of the sequences $\gamma_{\dot{A}_{\beta,\mu}}(h)$ and $\gamma_{\ddot{A}_{\beta,\mu}}(h)$, form the Toeplitz matrices $\Sigma(\dot{A}_{\beta,\mu})$ and $\Sigma(\ddot{A}_{\beta,\mu})$, and then compute the quadratic forms. Note that the inverse Fourier transforms of $\dot{A}_{\beta,\mu}$ and $\ddot{A}_{\beta,\mu}$ need only be determined once (see Section 2.3 for some explicit examples) and can be done ahead of time, and then applied repeatedly to many different time series.

In order to compute the time domain representation of the slope and convexity measures in (10), we utilize (8),

e.g., see the formulas in Examples 1 and 2. Of course, this will in general result in $\Sigma(\dot{A}_{\beta,\mu})$ and $\Sigma(\ddot{A}_{\beta,\mu})$ being complex. However, even if $\Sigma(g)$ (where g can be $A_{\beta,\mu}$, $\dot{A}_{\beta,\mu}$, or $\ddot{A}_{\beta,\mu}$) is a complex Toeplitz matrix, $X'\Sigma(g)X$ will always be real. From (8), it is easy to see that $\Sigma(g) = M + iN$ where M is real, symmetric, and Toeplitz, and N is real, skew-symmetric, and Toeplitz. Hence $X'NX = 0$ for any vector X, so that $X'\Sigma(g)X = X'MX$. Therefore, for the purposes of computing the statistical slope and convexity measures, we may take the real part of $\gamma_A(h)$ in (8).

Not only are these statistical estimates consistent, they are also asymptotically normal under some additional conditions (discussed in the Appendix). However, in order to construct a suitable normalization it will be necessary to estimate their variation. The asymptotic variance of $\theta_g(I)$ is $\theta_{g^2}(f^2)$ (if g is supported on $[0, \pi]$), which can be consistently estimated via $\theta_{g^2}(I^2)/2$. (The factor of 2 is required, since the integral of I^2 tends to the corresponding integral of $2f^2$ – see Chiu (1988)). This can be given a time-domain representation as follows. Let $R = \{R(1-n), ..., R(0), ..., R(n-1)\}'$ be a $2n-1$ vector of sample autocovariances, and let $\Sigma(g^2)$ be $2n-1$ dimensional in the following formula: $R'\Sigma(g^2)R/2 = \theta_{g^2}(I^2)/2$. This relationship can be easily verified using (9). Thus we will normalize $\theta_g(I)$ by the square root of $\theta_{g^2}(I^2)/2$. Hence our normalized statistical measures of slope and convexity are given by

$$-\psi_{\dot{A}_{\beta,\mu}}(I) = -\frac{\theta_{\dot{A}_{\beta,\mu}}(I)}{\sqrt{\theta_{\dot{A}^2_{\beta,\mu}}(I^2)/2}} = -\frac{1}{n}\frac{X'\Sigma(\dot{A}_{\beta,\mu})X}{\sqrt{R'\Sigma(\dot{A}^2_{\beta,\mu})R/2}}$$

and

$$\psi_{\ddot{A}_{\beta,\mu}}(I) = -\frac{\theta_{\ddot{A}_{\beta,\mu}}(I)}{\sqrt{\theta_{\ddot{A}^2_{\beta,\mu}}(I^2)/2}} = \frac{1}{n}\frac{X'\Sigma(\ddot{A}_{\beta,\mu})X}{\sqrt{R'\Sigma(\ddot{A}^2_{\beta,\mu})R/2}},$$

where the dimensions of the Σ matrices are either n or $2n-1$ as appropriate. The asymptotic properties of $\psi_{\dot{A}_{\beta,\mu}}(I)$ and $\psi_{\ddot{A}_{\beta,\mu}}(I)$ are discussed in the Appendix. In summary, both $-\sqrt{n}\psi_{\dot{A}_{\beta,\mu}}(I)$ and $\sqrt{n}\psi_{\ddot{A}_{\beta,\mu}}(I)$ are marginally asymptotically $N(0, 1)$ under $H_0^{(1)}$ and $H_0^{(2)}$ respectively and the assumptions discussed in the Appendix. Simulations indicate that the variance normalization is slow to converge, and its correlation with numerator causes a degree of non-normality in smaller samples. Based on the histogram of the distribution simulated under a Gaussian white noise Null hypothesis with $n = 360$ and 10,000 replications (Figure 1) there is close agreement to the normal distribution, except at the extremes in the tails. Section 4 explores this behavior further through simulation studies.

3.2 Applications to single peak testing

We now consider the application to peak testing. Recall that we have an initial Null Hypothesis $H_0^{(1)}$ that we must fail to reject in order to proceed. This can be interpreted as saying there is insufficient evidence to conclude that the first derivative (slope) of the spectral density is significantly different from zero. Now we know that $-\sqrt{n}\psi_{\dot{A}_{\beta,\mu}}(I)$ is asymptotically $N(0, 1)$ under $H_0^{(1)}$ and the assumptions discussed in the Appendix. If we further suppose that a sufficiently small value x is obtained for the test statistic, we will not be able to reject $H_0^{(1)}$ with any confidence. In that case, we can consider the hypothesis $H_0^{(2)}$, which we seek to reject; this is tested via $\sqrt{n}\psi_{\ddot{A}_{\beta,\mu}}(I)$. Although $-\sqrt{n}\psi_{\dot{A}_{\beta,\mu}}(I)$ and $\sqrt{n}\psi_{\ddot{A}_{\beta,\mu}}(I)$ are asymptotically correlated (see Theorem 1 of the Appendix) we will consider the slope and convexity tests as if they were done separately (this correlation can be estimated, and used to determine the distribution of the convexity diagnostic conditional on the slope diagnostic; however, the interpretation of p-values becomes muddled. For simplicity, we treat the tests separately, one at a time, and do not explicitly account for the correlation). Our testing procedure is then conducted as follows:

1. Perform the 2-sided test of $H_0^{(1)}$ using $-\sqrt{n}\psi_{\dot{A}_{\beta,\mu}}(I)$.
2. Let p be the p-value associated with the first test statistic's value $x = -\sqrt{n}\psi_{\dot{A}_{\beta,\mu}}(I)$, with x and p related by $p = 2\Phi(-|x|)$.
3. If $p > 0.05$ (or some other pre-determined tolerance level) proceed; else conclude that there is no peak present.
4. Perform the lower 1-sided test of $H_0^{(2)}$ using $\sqrt{n}\psi_{\ddot{A}_{\beta,\mu}}(I)$.
5. Reject $H_0^{(2)}$ and conclude that there is a peak if $\sqrt{n}\psi_{\ddot{A}_{\beta,\mu}}(I) < \Phi^{-1}(\alpha)$, where α is the level of the convexity test.

3.3 Joint peak testing: Application to seasonal adjustment

We now consider the situation where we wish to test for several spectral peaks simultaneously. Clearly we could design a kernel with several nodes, one at each peak, but this would merely be the sum of several individual spectral peak diagnostics. It would have the disadvantage that a significant spectral peak in one place could cancel a significant spectral trough elsewhere. Therefore, we would prefer a test that examines a set of spectral diagnostics within a multiple testing paradigm.

For example, consider the context of testing for spectral peaks in seasonally adjusted data. There are six seasonal peaks of interest, but we must restrict attention to five due to aliasing problems (the peak at frequency π cannot be identified). If one or more of the spectral peaks is significant, we must reject our seasonal adjustment procedure (since it has failed to remove all of the peaks); therefore, we are in a multiple testing situation, and will utilize a method that controls the familywise error rate (FWER) proposed by Hochberg (1988) and described in Benjamini and Hochberg (page 294, 1995). Restricting attention to the issue of convexity, we have Null Hypotheses $H_0^{(2)}$ for each of the five seasonal frequencies. In our setting, the procedure of Hochberg (1988) is to compute p-values for the convexity test at each of the five seasonal frequencies, and order them as $p_{(1)} \le p_{(2)} \le p_{(3)} \le p_{(4)} \le p_{(5)}$, with corresponding Null Hypotheses denoted by $H_{(i)}$. For a specified FWER of level α (*e.g.*, $\alpha = 0.05$), let k be the largest i for which $p_{(i)} \le i/(6 - i)\alpha$; then reject all $H_{(i)}$ for $i \le k$.

When using such a procedure, we should make Type I errors – *i.e.*, identifying at least one seasonal frequency as having negative convexity when none is present – roughly α proportion of the time (if we were to restrict attention to $H_0^{(2)}$, the convexity hypothesis). The advantage of the Hochberg familywise error rate approach (H-FWER) is that it dramatically improves the statistical power compared to other methods. The validity of this method requires independence of the test statistics under consideration, and so for this reason we take five kernels $A_1, ..., A_5$ – centered at the seasonal frequencies $\pi/6, ..., 5\pi/6$ respectively – that have disjoint support. Then Theorem 1 can be generalized to obtain asymptotic independence of the five convexity test statistics (see the discussion after Theorem 1 in the Appendix). Of course, we also conduct five separate tests of the slope at each seasonal frequency, where we must fail to reject in each case in order to proceed.

As a final remark, we note that in practice a seasonal adjustment is rarely rejected on the basis of significant spectral mass at the fifth seasonal frequency of $5\pi/6$ (Findley 2006). This is partly due to the difficulty in assigning an interpretation to this frequency. Therefore, practitioners may be more interested in a "four-peak test" that focuses on the first four seasonal frequencies; one obtains this test by an obvious modification of the H-FWER procedure described above.

3.4 Extending to nonstationary data

The methodology given above assumes that the data are a sample from a stationary process. However, in the context of seasonal adjustment, it is usually the case that the seasonally adjusted data are once or twice integrated. In this case one would difference the seasonally adjusted data once or twice, and then apply the diagnostics. Now application of the differencing operators $1 - B$ and $(1 - B)^2$ are essentially high-pass filters, which can be expected to attenuate residual spectral peaks close to frequency zero (in particular, the first seasonal frequency at $\pi/6$). Thus it may be desirable to apply the diagnostic to the pseudo-spectrum instead; this can be done if the support of the kernel is bounded away from the poles in the spectral density.

Suppose that the observed data are now $Y_{1-d}, ..., Y_n$ for d the order of trend differencing (so usually $d = 1$ or 2). When the observed data are differenced, we obtain the sample X, which is strictly stationary. The pseudo-spectral density of the $\{Y_t\}$ process is $g(\lambda) = f(\lambda)\,|1 - e^{-i\lambda}|^{-2d}$, where f is the spectrum of $\{X_t\}$. This pseudo-spectrum could be estimated via $\hat{g}(\lambda) = I(\lambda)\,|1 - e^{-i\lambda}|^{-2d}$, where I is the periodogram of X as before; this is the re-coloring approach of Nerlove (1964). Then $\theta_A(g)$ is well-defined so long as $A(\lambda)\,|1 - e^{-i\lambda}|^{-2d}$ is an integrable function; essentially we must ensure that frequency zero is excluded from the support of the kernel A. Since A is centered around seasonal frequencies in practice, we can easily contrive this condition. The corresponding estimator is then

$$\hat{\theta}_A(\ddot{g}) = \theta_{\ddot{A}b}(I),$$

where $b(\lambda) = |1 - e^{-i\lambda}|^{-2d}$. The estimator is well-defined if $\ddot{A}b$ is integrable; moreover the asymptotic properties discussed in the Appendix for the stationary case extend to this case as well, so long as $\ddot{A}b$ is bounded.

This extension may be more appealing to some researchers. However, the cost is that the inverse Fourier transform of $\ddot{A}b$ must be determined, which requires some additional mathematical work. In the simulation studies and data illustrations in Section 4 we trend difference the seasonally adjusted data, but do not implement the correction factor b in the kernel.

4. Empirical studies

Having developed the theoretical aspects of the spectral diagnostic, we now turn to its performance in practice. We first present some results obtained from simulation, which provide insight into the size and power properties of the test statistic in finite samples. Then we investigate the size and power empirically, by applying the spectral diagnostics to a suite of 130 time series (65 U.S. Census Bureau series and 65 OECD series); we consider both the original and the seasonally adjusted series, and make comparisons to the Visual Significance, M7 and M8 quality control diagnostics of X-12-ARIMA (U.S. Census Bureau 2002). Additional empirical studies can be found in Evans, Holan and McElroy (2006).

4.1 Simulation study

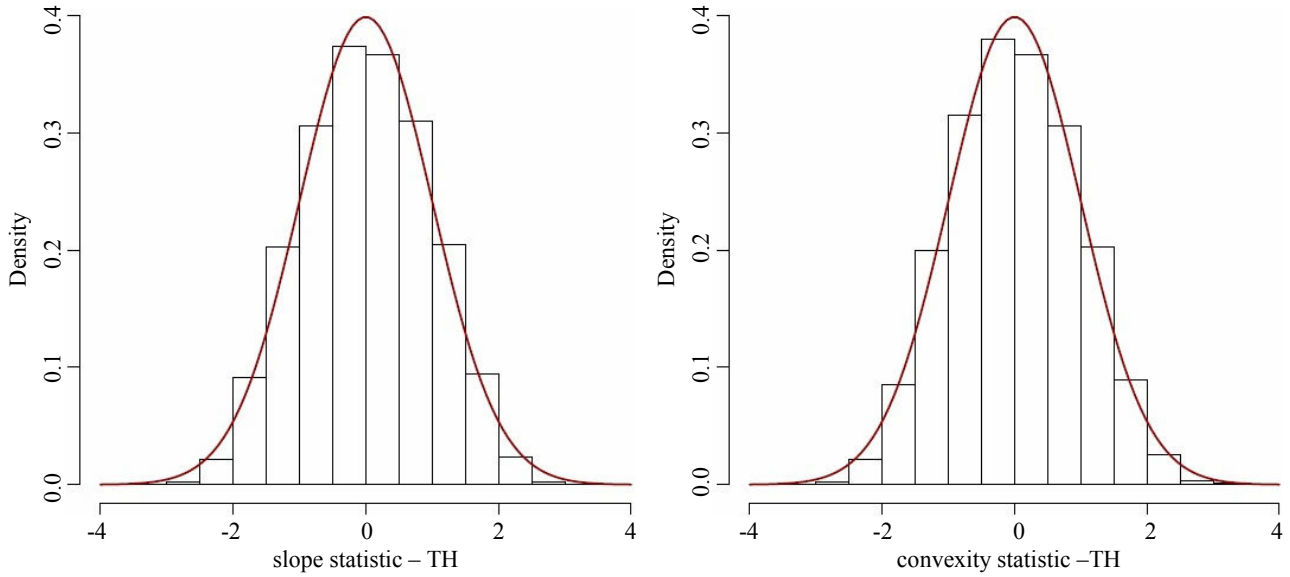
To evaluate the performance of our diagnostics we conducted several simulations. The first set of simulations examines size (level) for the single peak diagnostic. For this simulation we considered the slope and convexity diagnostics separately. Although in practice, when considering the slope diagnostic, we wish to fail to reject the Null hypothesis $H_0^{(1)}$, here we are interested in empirically investigating the distributional properties, and so we impose the usual definition of size for this study. So we simulated Gaussian white noise which satisfies the assumptions of Theorem 1 as well as satisfying $\psi_A(I) = \psi_{\dot{A}}(I) = 0$, so that $H_0^{(1)}$ and $H_0^{(2)}$ are true. Of course, there are many processes for which both $H_0^{(1)}$ and $H_0^{(2)}$ are true simultaneously – for example, any process with locally flat spectral density; however, due to asymptotic considerations it suffices to consider white noise. For a (large) sample size of $n = 360$, using the TH kernel with $\mu = \pi/6$ and $\beta = \pi/6$ (this corresponds to a kernel centered on the interval $[0, \pi/6]$), 10,000 repetitions yields an empirical distribution of the normalized diagnostics, $\psi_{\dot{A}}(I)$ and $\psi_{\ddot{A}}(I)$, whose histograms are displayed in Figure 1. Henceforth, let δ and α denote the levels associated with the slope and convexity tests respectively. Note that in this case we define level to mean the probability of rejecting $H_0^{(j)}$ $(j = 1, 2)$ when $H_0^{(j)}$ is true. Although, in practice, in the case of the slope hypothesis we wish to fail to reject we follow the strict definition of level and assume (for the purposes of this simulation) that the null hypothesis $H_0^{(1)}$ for the slope holds true. Similarly the null hypothesis for the convexity is $H_0^{(2)}$. Both the slope and convexity hypotheses are evaluated independently. Table 1 summarizes the results using both kernels from Section 2, for various sample sizes; the indicated δ, α-levels are for the nominal 5 percent level. Additionally, other choices of μ and β (not shown here) yielded similar results. As depicted in this study, in smaller samples, we observed skewness in the distribution which seems to be due to correlation between $\theta_A(I)$ and $\theta_{A^2}(I^2)$. Also, note that the size for the convexity test is larger for the quartic kernel than for the TH kernel.

Next we consider the empirical power for our single peak diagnostic. In this setting we evaluate the power based on a joint test of the slope and convexity. Specifically, we wish to fail to reject $H_0^{(1)}$ while simultaneously rejecting $H_0^{(2)}$, at $\delta = \alpha = 0.05$, and thus correctly identify spectral peaks. Since our composite Null hypothesis is that there is no peak, the Alternative hypothesis includes processes such as the $AR(2)$ given by

$$(1 - 2\rho\cos\omega B + \rho^2 B^2) X_t = \varepsilon_t \qquad (11)$$

with white noise variance σ^2, associated with some fixed frequency $\omega \in [0, \pi]$. The spectrum associated with the process in (11) is given by $f(\lambda) = \sigma^2 |1 - 2\rho\cos\omega e^{-i\lambda} + \rho^2 e^{-2i\lambda}|^{-2}$, which is maximized at $\lambda_0 = \cos^{-1}(\cos\omega(1 + \rho^2)/2\rho)$. Therefore one can explore the power of a peak-testing procedure by simulating from (11) with various choices of ρ, ω, and σ. Table 2 presents the result of 10,000 simulations, of various sample sizes, from the $AR(2)$ cycle model given in (11) with peak at $\mu = \pi/6$ and bandwidth set at $\beta = \pi/6$. The peak strength is parametrized through ρ, which we vary from 0.85 to 0.95; clearly $H_0^{(1)}$ and $H_a^{(2)}$ are both true for this model. In other words, there are spectral peaks, of different heights, at $\lambda = \pi/6$. This AR cycle model was chosen because it provides a convenient parametrization of spectral peak location and shape. Additionally, this choice of β is compatible with the seasonal adjustment setting, as this provides the maximum window width while avoiding overlapping spectral peaks. As expected, the power of our diagnostic increases with sample size and peakedness, ranging from 0.227 (quartic kernel) in small samples having weak spectral peak to ≈ 0.95 (TH kernel) in larger samples having a more pronounced spectral peak (see Table 2). Note that in this procedure the innovation variance is set equal to one, but it is immaterial due to the normalization of the diagnostic. In summary, both the quartic and TH kernels possess decent size and power properties. Generally, the quartic kernel seems to have superior size and power, so it would be preferable for spectra of this form (note that the lower power of the TH kernel is in part due to its being undersized). Additionally, it seems that smaller values of β (results not shown) require a greater sample size; a smaller β corresponds to a more refined "viewing" of the spectral peak, which would require more data to handle the resolution.

Although the individual peak testing scenario provides the foundation for our joint testing framework, as noted, the joint testing framework provides important methodology for applications in federal statistics. The application of importance is the evaluation of effective seasonal adjustment through the exploration of residual seasonality. Thus, it is of particular interest to know how our multiple testing approach performs in simulation. Therefore, in order to investigate the size and power associated with our joint test, we simulated 10,000 repetitions from a Gaussian white noise process and from an $AR(25)$ model obtained as a fit to the Current Employment Series (Employed Males, aged 16 to 19). Our goal in the power study was to construct an $AR(p)$ process (because of its ease in simulation and desirable theoretical properties as a parametric spectral estimator – see Parzen 1983) with (stationary) spectral peaks that are realistic, or close to what might be found in practice. Thus we obtain our $AR(25)$ model – fitted via maximum likelihood using AIC – which has similar seasonal dynamics (local spectral behavior) to the Current Employment Series (CES).



(a) (b)

Figure 1 Histogram of distribution of $-\sqrt{n}\,\psi_{\hat{A}_{\beta,\mu}}(I)$ (a) and $\sqrt{n}\,\psi_{\hat{A}_{\beta,\mu}}(I)$ (b) under a Gaussian white noise Null hypothesis using the TH kernel. The sample size is $n = 360$ with 10,000 replications

Table 1
Results of size simulation for the single peak diagnostic. Here $\mu = \beta = \pi/6$ and 10,000 repetitions were used. The slope and convexity diagnostics were investigated separately for both the quartic and TH kernels

Size for Single Peak $\mu = \beta = \pi/6$												
	Slope						Convexity					
	Quartic Kernel			TH Kernel			Quartic Kernel			TH Kernel		
n	Mean	Stdev	δ -level	Mean	Stdev	δ -level	Mean	Stdev	α -level	Mean	Stdev	α -level
120	0.003	0.903	0.007	-0.011	0.903	0.008	-0.065	0.852	0.032	0.025	0.888	0.018
144	-0.004	0.920	0.014	-0.011	0.927	0.015	-0.077	0.882	0.042	0.006	0.892	0.025
180	-0.003	0.942	0.022	0.002	0.920	0.017	-0.071	0.892	0.043	0.005	0.902	0.028
288	0.003	0.954	0.027	-0.002	0.950	0.025	-0.072	0.921	0.051	-0.006	0.926	0.033
360	0.003	0.962	0.032	-0.009	0.954	0.031	-0.056	0.922	0.051	0.006	0.951	0.040

Table 2
Results of power simulation for the single peak diagnostic. Here $\mu = \beta = \pi/6$ and 10,000 repetitions were used. The alternative hypothesis is given by the $AR(2)$ model defined by (11). The slope and convexity diagnostics were investigated simultaneously for both the quartic and TH kernels using $\delta = \alpha = 0.05$ for both tests (see Section 4.1)

Power for Single Peak $\mu = \beta = \pi/6$ - $(\delta, \alpha) = (0.05, 0.05)$						
	Quartic Kernel			TH Kernel		
n	$\rho = 0.85$	$\rho = 0.90$	$\rho = 0.95$	$\rho = 0.85$	$\rho = 0.90$	$\rho = 0.95$
120	0.227	0.438	0.758	0.147	0.335	0.670
144	0.287	0.532	0.856	0.208	0.431	0.799
180	0.354	0.643	0.923	0.272	0.567	0.901
288	0.447	0.755	0.949	0.372	0.706	0.950
360	0.601	0.872	0.937	0.537	0.859	0.948

To evaluate size we considered both a test based on convexity alone and a test based on the slope and convexity simultaneously. The tests based on convexity alone (C) were performed at the nominal α -levels of 0.05 and 0.10, using the H-FWER method to control the FWER. The tests based on both the slope and convexity simultaneously (S, C) were performed as follows:

1. Perform multiple tests of convexity, $H_0^{(2)}$, using the H-FWER method to control the FWER at level α (which is either 0.05 or 0.10).
2. For any peaks found significant in Step 1 perform individual slope test, $H_0^{(1)}$, at level δ (which is either 0.10 or 0.25). Note here we wish to fail to reject $H_0^{(1)}$ in order to declare any "peaks" as statistically significant.
3. Declare there is a statistically significant peak if Step 1 finds any seasonal frequency with significant aggregate convexity in the spectrum, and if Step 2 simultaneously fails to find any significant aggregate slope for the corresponding seasonal frequency.

The results of this simulation are summarized in Table 3. One aspect of this procedure that deserves further explanation is Step 2 where δ (the level for the slope test) is taken equal to 0.10 and 0.25. Although the slope testing aspect of the procedure is conducted on an individual peak basis, it seems reasonable to try and be conservative. The issue here is that even if some of the individual slope tests are rejected we may still proceed in other cases. Thus the situation encountered here differs from the classical "no peaks" hypothesis which can be rejected if a single peak is found. Of course, since we are conducting each slope hypothesis test on an individual peak basis, any δ -level greater than 0.05 would be considered more conservative.

While we cannot expect the combined procedure (S, C) to have size approaching the nominal (because using the slope test throws off the Type I error rate), neither is the size highly accurate in the case of using the convexity test alone (C), as can be seen by examining the $\alpha = 0.10$ case with $n = 288, 360$. Here the convexity for the quartic kernel is over-sized, whereas in the single peak case the convexity test has accurate size (Table 1) for these sample sizes. Note that H-FWER only produces an approximately correctly sized procedure; another factor is that the five peak tests are only asymptotically independent. It is for these reasons that the empirical size found in Table 3 differ somewhat from the nominal levels.

To investigate power we considered the same 3 step procedure as outlined above. However, for this simulation we only considered joint slope - convexity testing and examined four pairs of (δ, α) levels; $(\delta, \alpha) = (0.10, 0.05), (0.25, 0.05), (0.10, 0.10)$ and $(0.25, 0.10)$. The results of this simulation (Table 4) indicate tremendous power even at sample sizes as small as $n = 120$. This is extremely important as $n = 120$ is representative of the size samples encountered in practice when conducting seasonal adjustment (*i.e.*, 10 years of monthly data). For samples sizes $n = 144$, greater than 90% power was achieved.

Table 3
Results of size simulation for the multiple peak diagnostic. Here 10,000 repetitions were used. The convexity test was investigated separately with the FWER controlled at $\alpha = 0.05$ and $\alpha = 0.10$ using the H-FWER method. Additionally, the slope and convexity diagnostics were investigated simultaneously using the H-FWER method for the convexity controlling the FWER at $\alpha = 0.05$ and $\alpha = 0.10$, while the slope was evaluated at $\delta = 0.25$ and $\delta = 0.10$. Both the quartic and TH kernels were used for both tests (see Section 4.1)

	Size for Multiple Peaks H-FWER											
	C = 0.05		(S, C) = (0.10, 0.05)		(S, C) = (0.25, 0.05)		C = 0.10		(S, C) = (0.10, 0.10)		(S, C) = (0.25, 0.10)	
n	Quartic	TH	Quartic	TH	Quartic	TH	Quartic	TH	Quartic	TH	Quartic	TH
120	0.006	0.002	0.006	0.002	0.005	0.002	0.076	0.047	0.070	0.044	0.076	0.046
144	0.009	0.002	0.011	0.002	0.008	0.003	0.087	0.053	0.090	0.051	0.086	0.050
180	0.019	0.005	0.020	0.006	0.017	0.005	0.107	0.062	0.097	0.059	0.093	0.057
288	0.031	0.009	0.026	0.008	0.025	0.008	0.117	0.069	0.116	0.069	0.112	0.068
360	0.042	0.012	0.045	0.019	0.035	0.015	0.140	0.087	0.133	0.084	0.129	0.087

Table 4
Results of power simulation for the multiple peak diagnostic. Here 10,000 repetitions were used. The slope and convexity diagnostics were investigated simultaneously using the H-FWER method for the convexity controlling the FWER at $\alpha = 0.05$ and $\alpha = 0.10$. For the slope $\delta = 0.25$ and $\delta = 0.10$. Both the quartic and TH kernels were used for both tests (see Section 4.1)

	Power for Multiple Peaks H-FWER							
	(0.10, 0.05)		(0.25, 0.05)		(0.10, 0.10)		(0.25, 0.10)	
n	Quartic	TH	Quartic	TH	Quartic	TH	Quartic	TH
120	0.877	0.897	0.860	0.881	0.997	0.997	0.996	0.998
144	0.943	0.952	0.942	0.950	0.999	1.00	0.999	1.00
180	0.989	0.992	0.989	0.992	1.00	1.00	0.999	1.00
288	1.00	1.00	0.999	0.999	1.00	1.00	0.999	1.00
360	1.00	1.00	0.998	0.999	1.00	1.00	0.998	0.999

4.2 Case studies

We also considered 130 time series, 65 from the U.S. Census Bureau and 65 from OECD. These series consist of 35 U.S. Manufacturing series, 10 U.S. Housing series, 10 U.S. Import/Export series, and 10 U.S. Retail series; there are also 22 German series, 15 Euro-area series, 11 French series, and 17 Great Britain series from OECD, covering the sectors of manufacturing, retail, wholesale, foreign trade, unemployment, and industry. For every series, we computed the seasonal peak tests for both the raw data (logged) and the seasonally adjusted data (logged) – using the x11 specification of X-12-ARIMA – with both the quartic and TH kernels. We employed the H-FWER procedure controlling the FWER at $\alpha = 0.05$, 0.10, and where the threshold for the slope tests was $\delta = 0.25$ and $\delta = 0.10$ at each peak (see Section 4.1). Note that $\delta = 0.25$ and $\delta = 0.10$ produced similar results and thus, for the sake of brevity, only results for $\delta = 0.10$ are presented here. The results for $\delta = 0.25$ are available upon request from the first author. For both the raw and seasonally adjusted data, a single trend difference was used (as is the case for the Visual Significance diagnostic, described below) before applying the seasonal peaks test.

In addition, we present the M7 and M8 statistics as well as the results of the Visual Significance (VS) diagnostic both before and after adjustment. The M7 quality control statistic measures the amount of stable seasonality relative to the moving seasonality in the original series, with values greater than 1 indicating that the seasonality in the series is not identifiable (Lothian and Morry 1978); similarly, the M8 statistic measures the size of the fluctuations in the seasonal component, with a similar interpretation. We also considered the robust nonparametric Kruskal-Wallis test (U.S. Census Bureau 2002) for the presence of seasonality assuming stability. VS is based on an $AR(30)$ spectrum estimate of the raw and seasonally adjusted series, and is described in Soukup and Findley (1999). For Tables 5-10, each cell entry lists which seasonal frequencies were found to have a significant peak, with j corresponding to $\pi j/6$ for j = 1, 2, 3, 4, 5; an entry of $\emptyset$ indicates that no peaks were detected. For the M7 and M8 diagnostics, only the value is reported since there is no associated p-value (and they are only pertinent to the raw series).

The results of this empirical study can be found in Tables 5-10. All of the Kruskal-Wallis statistics were significant with $p = 0.000$, so these are not reported in the tables. The set of columns corresponding to the "Original Data" heading can be seen as giving empirical power (for each subset of series), assuming that each series is indeed seasonal and has seasonal spectral peaks. That is, the Total " correct " number gives the proportion of times each method correctly identified seasonality, and hence this proportion is a crude proxy for empirical power. We also report the average number of peaks that were identified, which is an empirical measure of the efficacy of the methods (the more peaks correctly identified, the better). The set of columns for "SA Data" gives an empirical size (for each subset of series), assuming that seasonal adjustment has indeed removed the spectral peaks. These are rough considerations, since we do not really know a priori whether the SA Data has been adequately adjusted.

The VS identifies all of the raw series as seasonal and most of the SA series as having no spectral seasonal peaks; the M7 and M8 diagnostics perform similarly, though of course they do not indicate which seasonal peaks are present in the raw data. Our procedure indicates a few cases (when $\alpha = 0.10$ for the convexity tests) where the adjustment may be inadequate, but these are within the scope of the expected proportion of Type I errors. For the raw series, the empirical power (*i.e.*, total proportion correct) for our method ranges between 0.66 and 0.89, with higher power for the $\alpha = 0.10$ level, as expected. In many cases the indicated peaks are the same as VS, but sometimes are quite different. Note that the average number of peaks detected for raw series was typically much higher for our procedure over the VS method, which often had an average around 3.2. When the α level was increased from 0.05 to 0.10, our method naturally increased in the average number of peaks detected; VS cannot be tuned in this way. Conversely, for SA data the average number of peaks detected tended to be less than one for our method (with the exception of the German series).

The results are fairly similar for the quartic and TH kernels. Although the M7, M8, and VS diagnostics have slightly better performance than our spectral peak procedure with $\alpha = 0.10$, it is important to note that our method provides a level of detail that M7 and M8 cannot replicate, while the VS diagnostic does not provide a p-value for any of the peaks (neither do M7 or M8). Overall, we find the results to be very encouraging and informative.

5. Conclusion

This paper presents an innovative approach to the statistical identification of spectral peaks. The convexity diagnostic computes an average of the periodogram weighted by the second derivative of a typical kernel, such as the Tukey-Hanning lag window. Implicitly this type of statistic involves a comparison of an average of the periodogram near a given frequency to its average somewhat further out; this follows from the general shape of $\dot{A}_{\beta,\mu}$. The slope diagnostic helps to screen out cases where there is negative convexity but also a large increase/decrease in the spectrum. That the method actually works as intended is borne out by the simulations and analysis results reported in Tables 1-10.

Table 5
Data analyses for 35 Manufacturing Series (U.S. Census Bureau) comparing our multiple peak diagnostic with VS, M7, and M8 diagnostics. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = 0.05$ and $\alpha = 0.10$ and examines the slope at $\delta = 0.10$ (see Section 4.2)

	Data Analyzes – Manufacturing Series											
	Original Data							SA Data				
	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS	M7	M8	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS
series	quartic	TH	quartic	TH				quartic	TH	quartic	TH	
M_1	2345	1235	2345	1235	12	0.24	0.39	ø	ø	ø	ø	ø
M_2	12345	12345	12345	12345	1235	0.20	0.32	ø	ø	ø	ø	ø
M_3	12345	12345	12345	12345	1235	0.28	0.46	ø	ø	ø	ø	ø
M_4	ø	ø	12345	ø	12	0.28	0.44	ø	ø	ø	ø	ø
M_5	12345	12345	12345	12345	12345	0.27	0.47	ø	ø	12345	12345	ø
M_6	ø	ø	123	123	12	0.28	0.49	ø	ø	ø	ø	ø
M_7	ø	ø	123	123	24	0.50	0.79	ø	ø	ø	ø	ø
M_8	12345	12345	12345	12345	12345	0.18	0.37	ø	ø	ø	ø	ø
M_9	12345	12345	12345	12345	124	0.42	0.73	ø	ø	ø	ø	ø
M_{10}	ø	ø	ø	ø	1	0.38	0.72	ø	ø	ø	ø	ø
M_{11}	ø	ø	12345	1234	123	0.15	0.27	ø	ø	ø	ø	ø
M_{12}	1234	1234	12345	12345	1234	0.30	0.54	ø	ø	ø	ø	ø
M_{14}	ø	ø	1234	1234	1234	0.24	0.39	ø	ø	ø	ø	ø
M_{15}	12345	12345	12345	12345	12345	0.23	0.43	ø	ø	ø	ø	ø
M_{16}	1234	1234	1234	1234	1234	0.23	0.40	ø	ø	ø	ø	ø
M_{17}	ø	ø	1234	12345	12	0.64	0.66	ø	ø	ø	ø	ø
M_{18}	12345	12345	12345	12345	245	0.20	0.37	ø	ø	ø	ø	ø
M_{19}	ø	ø	ø	ø	4	0.86	1.00	ø	ø	ø	ø	ø
M_{20}	ø	ø	ø	12345	4	0.56	0.84	ø	ø	ø	ø	ø
M_{21}	12345	12345	12345	12345	1234	0.37	0.58	ø	ø	ø	ø	ø
M_{22}	12345	12345	12345	12345	1234	0.26	0.45	ø	ø	ø	ø	ø
M_{23}	12345	12345	12345	12345	1234	0.20	0.47	ø	ø	ø	ø	ø
M_{24}	12345	12345	12345	12345	2345	0.26	0.43	ø	ø	ø	ø	ø
M_{25}	12345	12345	12345	12345	12345	0.27	0.42	ø	ø	ø	ø	ø
M_{26}	12345	12345	12345	12345	1235	0.37	0.62	ø	ø	ø	ø	ø
M_{27}	1345	1234	1345	1234	2345	0.25	0.22	ø	ø	ø	ø	ø
M_{28}	ø	ø	ø	ø	24	0.57	0.44	ø	ø	ø	ø	ø
M_{29}	ø	12345	12345	12345	24	0.78	1.13	ø	ø	ø	ø	ø
M_{30}	123	1234	12345	12345	245	0.45	0.65	ø	ø	ø	ø	ø
M_{31}	ø	ø	123	123	4	0.64	0.46	ø	ø	1234	1234	ø
M_{32}	1235	12345	1235	12345	12345	0.21	0.37	ø	ø	ø	ø	ø
M_{33}	12345	12345	12345	1234	1234	0.24	0.38	ø	ø	ø	ø	ø
M_{34}	12345	12345	12345	12345	234	0.46	0.85	ø	ø	ø	ø	ø
M_{35}	12345	12345	12345	12345	2345	0.25	0.66	ø	ø	ø	ø	ø
M_{36}	12345	12345	12345	12345	123	1.32	1.56	ø	ø	ø	ø	ø
Total "correct"	23/35	24/35	31/35	31/35	35/35	34/35	33/35	35/35	35/35	33/35	33/35	35/35
Average Number	3.09	3.29	4.09	4.09	3.23			0	0	0.26	0.26	0

Table 6
Data analyses for 30 U.S Census Bureau Series (10 Housing, 10 Import/Export and 10 Retail Sales) comparing our multiple peak diagnostic with VS, M7, and M8 diagnostics. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = 0.05$ and $\alpha = 0.10$ and examines the slope at $\delta = 0.10$ (see Section 4.2)

	Data Analyzes – Manufacturing Series											
	Original Data							SA Data				
	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS	M7	M8	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS
series	quartic	TH	quartic	TH				quartic	TH	quartic	TH	
MW1Fam	12345	12345	12345	12345	12	0.13	0.25	ø	ø	ø	ø	ø
NWTot	12345	12345	12345	12345	12	0.18	0.31	ø	ø	ø	ø	ø
NE1Fam	12345	12345	12345	12345	12	0.16	0.33	ø	ø	ø	ø	ø
NETot	12345	12345	12345	12345	123	0.25	0.27	ø	ø	ø	ø	ø
S1Fam	12345	12345	12345	12345	125	0.22	0.47	ø	ø	ø	ø	ø
STot	124	124	1245	1245	125	0.29	0.57	ø	ø	ø	ø	ø
US1Fam	12345	12345	12345	12345	125	0.17	0.39	ø	ø	ø	ø	ø
USTot	12345	12345	12345	12345	125	0.20	0.42	ø	ø	ø	ø	ø
W1Fam	1234	1234	12345	12345	125	0.21	0.44	ø	ø	ø	ø	ø
WTot	1234	1234	12345	1234	12	0.27	0.56	ø	ø	ø	ø	ø
Total "correct"	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
	Import/Export Series											
	Original Data							SA Data				
	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS	M7	M8	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS
series	quartic	TH	quartic	TH				quartic	TH	quartic	TH	
M00120	12345	ø	12345	12345	125	0.23	0.48	ø	ø	ø	ø	ø
M00190	12345	12345	12345	12345	1235	0.38	0.59	ø	ø	ø	ø	ø
M3000	12345	12345	12345	12345	234	0.48	0.95	ø	ø	ø	ø	ø
M3010	1234	1234	1234	12345	2345	0.52	0.88	ø	ø	ø	ø	ø
M12060	12345	12345	12345	12345	123	0.53	0.77	ø	ø	ø	ø	ø
X3	12345	12345	12345	12345	2345	0.57	0.94	ø	ø	ø	ø	ø
X00300	134	134	134	134	2	0.56	0.97	ø	ø	ø	ø	ø
X3020	12345	12345	12345	12345	12345	0.39	0.70	ø	ø	ø	ø	ø
X3022	12345	12345	12345	12345	23	0.69	1.04	ø	ø	ø	ø	ø
X10140	1234	1234	1234	1234	15	0.29	0.47	ø	ø	ø	ø	ø
Total "correct"	10/10	9/10	10/10	10/10	10/10	10/10	9/10	10/10	10/10	10/10	10/10	10/10
	Retail Series											
	Original Data							SA Data				
	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS	M7	M8	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS
series	quartic	TH	quartic	TH				quartic	TH	quartic	TH	
s0b441x0	12345	12345	12345	12345	135	0.22	0.41	ø	ø	ø	ø	ø
s0b 44000	12345	12345	12345	12345	2345	0.12	0.26	ø	ø	ø	ø	ø
s0b 44100	12345	12345	12345	12345	135	0.21	0.40	ø	ø	ø	ø	ø
s0b 44130	12345	12345	12345	12345	1235	0.21	0.42	ø	ø	ø	ø	ø
s0b 44200	12345	12345	12345	12345	12345	0.13	0.27	ø	ø	ø	ø	ø
s0b 44300	1234	12345	1234	12345	12345	0.12	0.18	ø	ø	ø	ø	ø
s0b 44312	1234	1234	1234	1234	12345	0.31	0.48	ø	ø	ø	ø	ø
s0b 44400	12345	12345	12345	12345	1235	0.16	0.32	ø	ø	ø	ø	ø
s0b 44410	12345	12345	12345	12345	1235	0.14	0.32	ø	ø	ø	ø	ø
s0b 44500	12345	12345	12345	12345	235	0.14	0.23	ø	ø	ø	ø	ø
Total "correct"	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Grand Total "correct"	30/30	29/30	30/30	30/30	30/30	30/30	29/30	30/30	30/30	30/30	30/30	30/30
Average Number	4.67	4.5	4.77	4.8	3.23			0	0	0	0	0

Table 7
Data analyses for 22 German OECD Series comparing our multiple peak diagnostic with VS, M7, and M8 diagnostics. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = 0.05$ and $\alpha = 0.10$ and examines the slope at $\delta = 0.10$ (see Section 4.2)

Data Analyzes – OECD DEU												
	Original Data							SA Data				
Series	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS	M7	M8	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS
DEU	quartic	TH	quartic	TH				quartic	TH	quartic	TH	
PRMNCG03	12345	12345	12345	12345	12345	0.17	0.33	ø	ø	ø	ø	ø
PRMNCS01	1234	1234	1234	1234	235	0.25	0.79	ø	ø	ø	ø	ø
PRMNIG01	12345	12345	12345	12345	12345	0.28	0.48	ø	ø	ø	ø	ø
PRMNTO01	134	134	134	134	12345	0.26	0.46	ø	ø	ø	ø	ø
PRMNVG01	1235	1235	1235	1235	2345	0.29	0.46	ø	ø	ø	ø	ø
SLMNCD01	1245	1235	1245	1235	23	0.22	0.40	ø	ø	ø	ø	ø
SLMNCN01	1345	2345	1345	2345	123	0.37	0.72	ø	ø	ø	ø	ø
SLMNDM01	2345	2345	2345	2345	123	0.32	0.63	ø	ø	ø	ø	ø
SLMNEX01	12345	12345	12345	12345	12	0.32	0.51	1:5	ø	12345	12345	ø
SLMNIG01	2345	2345	2345	2345	123	0.21	0.65	ø	ø	ø	ø	ø
SLMNTO01	245	345	245	345	23	0.20	0.66	ø	ø	234	ø	ø
SLRTCR01	1345	1345	1345	1345	1234	0.19	0.51	ø	ø	ø	ø	ø
SLRTTO01	12345	12345	12345	12345	12345	0.12	0.25	ø	ø	ø	ø	ø
SLRTTO02	12345	12345	12345	12345	12345	0.13	0.29	ø	ø	ø	ø	ø
SLWHTO01	2345	2345	2345	2345	123	0.20	0.62	ø	ø	134	123	ø
SLWHTO02	2345	2345	2345	2345	123	0.20	0.62	ø	ø	134	123	ø
UNLVRG01	23	23	23	23	124	0.23	0.48	ø	ø	1	ø	5
UNLVSUMA	345	345	345	345	12	0.30	0.53	12	ø	1245	12345	ø
UNLVSUTT	234	234	234	234	12	0.24	0.53	ø	ø	45	45	25
UNRTRG01	235	1245	235	1245	123	0.19	0.59	ø	ø	2345	2345	ø
XTEXVA01	1234	1234	1234	1234	23	0.28	0.77	ø	ø	ø	ø	ø
XTIMVA01	234	1234	2345	12345	23	0.31	0.95	ø	ø	ø	ø	ø
Total "correct"	22/22	22/22	22/22	22/22	22/22	22/22	22/22	20/22	22/22	14/22	16/22	20/22
Average Number	3.86	3.95	3.91	4.00	3.23			0	0	1.14	1.00	0.14

Table 8
Data analyses for 15 Euro-area OECD Series comparing our multiple peak diagnostic with VS, M7, and M8 diagnostics. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = 0.05$ and $\alpha = 0.10$ and examines the slope at $\delta = 0.10$ (see Section 4.2)

Data Analyzes – OECD EMU												
	Original Data							SA Data				
series	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS	M7	M8	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS
EMU	quartic	TH	quartic	TH				quartic	TH	quartic	TH	
PRCNTO01	345	1345	345	1345	12345	0.14	0.44	ø	ø	ø	ø	ø
PRINTO01	12345	12345	12345	12345	12345	0.10	0.23	ø	ø	ø	ø	ø
PRMNCG03	1245	1245	1245	1245	12345	0.12	0.32	1	1	1234	1234	ø
PRMNCS01	1234	12345	1234	12345	1234	0.22	0.47	ø	ø	ø	ø	ø
PRMNIG01	12345	12345	12345	12345	2345	0.15	0.27	ø	ø	ø	ø	ø
PRMNTO01	2345	2345	2345	2345	2345	0.14	0.23	ø	ø	1	ø	ø
PRMNVG01	1234	12345	1234	12345	2345	0.13	0.23	ø	ø	ø	ø	ø
SLMNCN02	12345	12345	12345	12345	123	0.31	0.57	ø	ø	ø	ø	ø
SLMNIG02	12345	12345	12345	12345	23	0.21	0.45	ø	ø	ø	ø	ø
SLMNTO02	1345	2345	1345	2345	23	0.20	0.41	24	ø	24	34	ø
SLMNVG02	12345	12345	12345	12345	2345	0.17	0.30	1	1	1245	1345	ø
SLRTTO01	12345	12345	12345	12345	12345	0.05	0.12	ø	ø	ø	ø	ø
SLRTTO02	12345	12345	12345	12345	12345	0.05	0.11	ø	ø	ø	ø	ø
XTEXVA01	1345	2345	1345	2345	23	0.31	0.57	ø	ø	12	12	ø
XTIMVA01	2345	2345	2345	2345	23	0.40	0.72	ø	ø	ø	ø	ø
Total "correct"	15/15	15/15	15/15	15/15	15/15	15/15	15/15	12/15	13/15	10/15	11/15	15/15
Average Number	4.40	4.60	4.40	4.60	3.73			0.20	0.13	0.87	0.80	0

Table 9
Data analyses for 11 French OECD Series comparing our multiple peak diagnostic with VS, M7, and M8 diagnostics. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = 0.05$ and $\alpha = 0.10$ and examines the slope at $\delta = 0.10$ (see Section 4.2)

Data Analyzes – OECD FRA												
	Original Data							SA Data				
series	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS	M7	M8	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS
FRA	quartic	TH	quartic	TH				quartic	TH	quartic	TH	
PRAFAG01	12345	12345	12345	12345	123	0.13	0.29	ø	ø	ø	ø	ø
PRNCTO01	2345	2345	2345	2345	235	0.14	0.44	ø	ø	ø	ø	ø
PRMNCG01	12345	12345	12345	12345	234	0.15	0.38	1	1	1	1	ø
PRMNCS01	12345	12345	12345	12345	234	0.25	0.58	ø	ø	ø	ø	ø
PRMNIG01	12345	12345	12345	12345	12345	0.11	0.26	ø	ø	ø	ø	ø
PRMNTO01	12345	12345	12345	12345	12345	0.16	0.29	123	123	123	1234	ø
PRMNVE01	12345	12345	12345	12345	12345	0.24	0.34	ø	ø	1245	1245	ø
SLRTCR01	1345	2345	1345	2345	123	0.27	0.71	ø	ø	ø	ø	ø
SLRTTO02	12345	12345	12345	12345	12345	0.16	0.36	ø	ø	ø	ø	ø
XTEXVA01	1345	1345	1345	1345	23	0.14	0.44	ø	ø	ø	ø	ø
XTIMVA01	1245	1245	1245	1245	23	0.18	0.54	ø	ø	ø	ø	ø
Total "correct"	11/11	11/11	11/11	11/11	11/11	11/11	11/11	9/11	9/11	8/11	8/11	11/11
Average Number	4.64	4.64	4.64	4.64	3.55			0.36	0.36	0.73	0.81	0

Table 10
Data analyses for 17 Great Britain OECD Series comparing our multiple peak diagnostic with VS, M7, and M8 diagnostics. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = 0.05$ and $\alpha = 0.10$ and examines the slope at $\delta = 0.10$ (see Section 4.2)

Data Analyzes – OECD GBR												
	Original Data							SA Data				
series	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS	M7	M8	H-FWER 0.10/0.05		H-FWER 0.10/0.10		VS
GBR	quartic	TH	quartic	TH				quartic	TH	quartic	TH	
PPIAMP01	1234	1234	1234	1234	24	0.56	1.58	ø	ø	ø	ø	ø
PPIAMP02	1	1	123	123	2	0.53	0.92	ø	ø	ø	ø	ø
PPIPFU01	1345	12345	1345	12345	12	0.64	0.59	ø	ø	ø	ø	ø
PRINTO01	1345	1345	1345	1345	23	0.16	0.40	ø	ø	ø	ø	ø
PRMNCG02	2345	2345	2345	2345	123	0.23	0.56	ø	ø	ø	ø	ø
PRMNCG03	12345	12345	12345	12345	123	0.20	0.49	ø	ø	ø	ø	ø
PRMNCS01	123	12	123	1234	12	0.68	1.31	ø	ø	ø	ø	ø
PRMNIG01	2345	2345	2345	2345	123	0.15	0.47	1	ø	ø	ø	ø
PRMNTO01	1345	1345	1345	1345	23	0.17	0.45	ø	ø	ø	ø	ø
PRMNVE02	12345	12345	12345	12345	12345	0.25	0.76	ø	ø	ø	ø	ø
PRMNVE03	1234	1234	1234	1234	234	0.29	0.91	ø	ø	ø	ø	ø
PRMNVG01	124	134	124	134	234	0.18	0.58	ø	ø	ø	ø	ø
SLRTCR03	1345	1345	1345	1345	124	0.42	0.74	124	123	124	123	ø
SLRTTO02	12345	12345	12345	12345	12345	0.05	0.15	1234	ø	1234	12345	ø
UNLVRG01	12345	12345	12345	12345	1245	0.63	0.61	ø	ø	ø	ø	ø
XTEXVA01	134	134	1345	1345	23	0.34	1.02	ø	ø	ø	ø	ø
XTIMVA01	23	23	23	23	23	0.31	0.90	ø	ø	ø	ø	ø
Total "correct"	17/17	17/17	17/17	17/17	17/17	17/17	14/17	14/17	16/17	13/17	15/17	17/17
Average Number	3.76	3.76	3.94	4.06	2.76			0.47	0.18	0.53	0.47	0

For the multiple peak-testing scenario, we employ known results from the multiple testing literature (*i.e.*, the applications to controlling FWER) to combine the p-values from the five seasonal frequencies in such a way to dramatically increase statistical power, as demonstrated in Table 4. Although there is some departure in the size (Table 3) for the multiple peak testing, the results are still quite usable. On a typical batch of seasonal series, the number of Type I errors are as expected, and the power is quite decent (Tables 5-10). While our method compares quite favorably to the VS, M7, and M8 diagnostics, neither of the diagnostics provide a p-value and only the former can distinguish which spectral peaks are contributing to seasonal behavior. This aspect is important to the seasonal adjuster, who wants to know not only that there may be residual seasonality, but also at what seasonal frequencies, so as to take appropriate action to alter the seasonal adjustment filters (this can be done by smoothing over additional years, which is accomplished by changing the seasonal filters in X-11-ARIMA; alternatively, one might consider shortening the series. For current research on a model-based approach to designing SA filters targeted for specific seasonal frequencies, see Aston, Findley, McElroy, Wills, and Martin (2007)).

The choice of kernel surely has some impact on the results, although we found little difference in practice between the quartic and TH kernels; the TH may be marginally more powerful. Of course, plenty of other popular kernels may also be utilized by a practitioner, and we have only chosen two that seemed intuitive and straightforward to implement. The choice of the location μ is clearly dictated by the characterization of seasonality. Since statistical power generally decreased with β, we always recommend taking the maximal β such that the kernel supports are disjoint, which guarantees the asymptotic independence property of the various diagnostics that is crucial to our multiple testing method.

Finally, the asymptotic results require that the data be differenced to stationarity. Recognizing that economic time series are typically nonstationary, it is desirable to trend-difference seasonally adjusted data before applying our diagnostic. This differencing may dampen the detection of the first seasonal peak, so practitioners may "re-color" the data as described in Section 3.4.

Acknowledgements

This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. Holan's research was supported by an ASA/NSF/BLS research fellowship.

Appendix

Here we derive asymptotic formulas for the statistical measures ψ_A of slope and convexity. These results can then be applied in the testing paradigm to get asymptotic critical values. Some mild conditions on the data are required for the asymptotic theory; we follow the material in Taniguchi and Kakizawa (2000, Section 3.1.1). Condition (B), due to Brillinger (1981), states that the process is strictly stationary and condition (B1) of Taniguchi and Kakizawa (2000, page 55) holds. Condition (HT), due to Hosoya and Taniguchi (1982), states that the process has a linear representation, and conditions (H1) through (H6) of Taniguchi and Kakizawa (2000, pages 55-56) hold. Assumption 1 (8) of Chiu (1988) is a summability condition on various higher order cumulants, which is satisfied, for example, by a Gaussian process with spectral density in C^2. None of these conditions are stringent; for example, a causal linear process with fourth moments satisfies (HT). The main result is a joint convergence of any two measures $\psi_A(I)$; *e.g.*, these can be a slope and convexity measure with the same kernel A. We present the general theorem that covers these two cases.

Theorem 1 *Suppose that the fourth order cumulants of* $\{X_t\}$ *vanish*; *that either condition* (*B*) *or* (*HT*) *holds*; *and that Assumption* 1 (8) *of Chiu* (1988) *holds*. *Let the kernels* A *and* B *satisfy conditions* (*i*) *through* (*iv*) *of Section* 2.1. *Then*

$$\left\{ \sqrt{n}\,\frac{(\theta_A(I) - \theta_A(f))}{\sqrt{\theta_{A^2}(I^2)/2}}, \right.$$

$$\left. \sqrt{n}\,\frac{(\theta_B(I) - \theta_B(f))}{\sqrt{\theta_{B^2}(I^2)/2}} \right\} \stackrel{\mathcal{L}}{\Rightarrow} N(0, V)$$

as $n \to \infty$. *Here* 0 *denotes the zero vector* $(0, 0)'$, *and* V *is a* 2×2 *matrix with entries*

$$V_{11} = V_{22} = 1 \quad V_{12} = V_{21} = \frac{\theta_{AB}(f^2)}{\sqrt{\theta_{A^2}(f^2)\theta_{B^2}(f^2)}}.$$

Proof. First we establish that $\theta_{A^2}(I^2) \xrightarrow{a.s.} 2\theta_{A^2}(f^2)$. Since the kernel A is continuous in an interval (such as $[\mu - \beta/2, \mu + \beta/2]$), this result follows directly from Corollary 1 of Chiu (1988), noting that they deal with the Riemann sums approximation to the integral functional (Chiu (1988) also defines the periodogram with a 2π factor). Of course the same results holds with B in place of A. Secondly, consider the joint convergence of $\theta_A(I)$ and

$\theta_B(I)$. We use the Cramer-Wold device, and apply Lemma 3.1.1 of Taniguchi and Kakizawa (2000), appropriately generalized to include non-even functions (*cf.* Theorem 3 of Chiu (1988)). Hence for any x, y real,

$$\sqrt{n}\left(x\,\frac{(\theta_A(I) - \theta_A(f))}{\sqrt{\theta_{A^2}(f^2)}} + y\,\frac{(\theta_B(I) - \theta_B(f))}{\sqrt{\theta_{B^2}(f^2)}} \right)$$

$$\overset{\mathcal{L}}{\Rightarrow} N\left(0,\ \frac{1}{2\pi}\int_{-\pi}^{\pi}(C(\lambda)C(-\lambda) + C^2(\lambda))\,f^2(\lambda)\,d\lambda\right)$$

using Slutsky's Theorem (Bickel and Doksum 1977), where the kernel C is defined by

$$C(\lambda) = \frac{x}{\sqrt{\theta_{A^2}(f^2)}}\,A(\lambda) + \frac{y}{\sqrt{\theta_{B^2}(f^2)}}\,B(\lambda).$$

Clearly $C(\lambda)\,C(-\lambda) = 0$ and

$$C^2(\lambda) = \frac{x^2}{\theta_{A^2}(f^2)}\,A^2(\lambda) + 2\frac{xy}{\sqrt{\theta_{A^2}(f^2)}\sqrt{\theta_{B^2}(f^2)}}\,A(\lambda)\,B(\lambda) + \frac{y^2}{\theta_{B^2}(f^2)}\,B^2(\lambda).$$

By taking x and y to be zero and one in various combinations, we deduce the stated variance matrix V.

Next we discuss the multiple-peak testing scenario. So suppose that we have a finite collection of kernels A_i for $i = 1, 2, \ldots, d$, each of which satisfies the assumptions of Section 2. Then we can easily generalize Theorem 1 from two to d kernels as follows. The asymptotic covariance matrix V will have ij^{th} entry

$$\frac{\theta_{A_iA_j}(f^2)}{\sqrt{\theta_{A_i^2}(f^2)\theta_{A_j^2}(f^2)}}.$$

Thus, if the support for any two kernels is disjoint we obtain asymptotic independence, and can therefore invoke the H-FWER multiple testing procedure.

References

Aston, J., Findley, D., McElroy, T., Wills, K. and Martin, D. (2007). New ARIMA Models for Seasonal Time Series and Their Application to Seasonal Adjustment and Forecasting. *SRD Research Report No. RRS 2007*-14, *U.S. Census Bureau.*

Bell, W., and Hillmer, S. (1984). Issues involved with the seasonal adjustment of economic time series. *Journal of Business and Economic Statistics*, 2, 291-320.

Bickel, P., and Doksum, K. (1977). *Mathematical Statistics*: *Basic Ideas and Selected Topics*. Englewood Cliffs, New Jersey: Prentice Hall.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, Series B, 57, 289-300.

Brillinger, D. (1981). *Time Series Data Analysis and Theory*. San Francisco: Holden-Day.

Chiu, S. (1988). Weighted least squares estimators on the frequency domain for the parameters of a time series. *The Annals of Statistics*, 16, 1315-1326.

Evans, T., Holan, S. and McElroy, T. (2006). Evaluating measures for assessing spectral peaks. *2006 Proceedings American Statistical Association*, [CD-ROM]: Alexandria, VA}.

Findley, D. (2006). Personal communication.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. and Chen, B.C. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business and Economic Statistics*, 16, 127-177 (with discussion).

Hosoya, Y., and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, 10, 132-153.

Lothian, J., and Morry, M. (1978). A test for identifiable seasonality when using the X-11-ARIMA program. Working Paper, Time Series Research and Analysis Division, Statistics Canada.

Maravall, A., and Caporello, G. (2004). Program TSW: Revised Reference Manual. *Working Paper 2004*, *Research Department*, *Bank of Spain*. http://www.bde.es.

Nerlove, M. (1964). Spectral analysis of seasonal adjustment procedures. *Econometrica*, 32, 241-286.

Newton, H., and Pagano, M. (1983). A method for determining periods in time series. *Journal of the American Statistical Association*, 78, 152-157.

Parzen, E. (1983). Autoregressive spectral estimation. *Handbook of Statistics III*, (Ed. D. Brillinger and P. Krishnaiah), Amsterdam: North Holland, 221-247.

Priestley, M. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

Soukup, R.J., and Findley, D.F. (1999). On the spectrum diagnostics used by X-12-ARIMA to indicate the presence of trading day effects after modeling or adjustment. Also www.census.gov/pub/ts/papers/rr9903s.pdf. *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 144-149.

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

U.S. Census Bureau (2002). X-12 ARIMA Reference Manual (Version 0.2.10), Washington, DC.