Using Spectral Peaks to Detect Seasonality

Tucker McElroy* and Scott Holan†
U.S. Census Bureau and University of Missouri-Columbia

Abstract

Peaks in the spectrum of a stationary time series indicate the presence of periodic phenomena, such as seasonal and cyclical components. Extending the approach of McElroy and Holan (2009), this paper constructs nonparametric peak-estimators via windowing the periodogram with the same sinusoidal function used to define seasonal autocovariances. We also utilize a multiple testing procedure that controls the familywise error rate, as in Hochberg (1988) and McElroy and Holan (2009). The application of these methods is exceedingly simple, requiring only a few lines of computer code; therefore the method can be used to quickly evaluate series for residual seasonality. Simulation studies demonstrate the good size and power of the procedure, and an extensive data study further illustrates the method.

Keywords. Seasonal adjustment; Spectral density; Time series.

1 Introduction

The detection of peaks in the spectrum of a time series has several important applications. Peaks are indicative of periodic or quasi-periodic behavior in the time series, and can be associated with a wide range of phenomena: trends, cycles, seasonality, or trading day effects. In this paper we focus on seasonality, where the peak locations are always known ahead of time, but there remains a question of whether the spectrum has significant mass at these seasonal frequencies. In the context of a federal agency such as the U.S. Census Bureau that seasonally adjusts thousands of economic time series each month, the statistician engaged in production is interested in having a quick and powerful tool to assist with handling seasonality nonparametrically. In particular one wants to know: (1) is the original (or raw) series seasonal; (2) does the output of my seasonal adjustment procedure have residual seasonality?

With regard to the first question, only series exhibiting seasonality should be seasonally adjusted. Given the sensitivity and importance of published series, the decision to seasonally adjust should not be made subjectively. Current practice at the U.S. Census Bureau often utilizes examination of spectral plots and the use of Visual Significance (VS) to gauge seasonality of the raw data – see the discussion in Findley, Monsell, Bell, Otto, and Chen (1998). As for the second question above, if the series has just been adjusted with some method (such as X-12 ARIMA or TRAMO-SEATS) the statistician wants to know if the seasonality has been adequately removed – otherwise it may be necessary to do some additional modeling and refinement of the inputs to the software. Common practice at the U.S. Census Bureau is again to use VS and spectral plots to directly assess this. The proposed method of this paper aims to improve upon the non-statistical VS procedure.

Earlier work on this topic can be found in McElroy and Holan (2009), which is broader in some ways than the present work. As noted, VS uses the spectrum but does not associate a statistical significance, or

*Statistical Research Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov
†Department of Statistics,University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO, 65211-6100, holans@missouri.edu

p-value, to the resulting decision problem; the spectral diagnostics of McElroy and Holan (2009) – referred to as MH tests – rectify this using measures of slope and convexity. The paper at hand simplifies this to a single measure (for each peak), thus simplifying the distribution theory. Instead of having two tests to measure the slope and convexity of a peak, we now use a single measure that essentially computes a weighted spectral mass – and is very intuitively related to the sample autocovariance estimate at seasonal lags. Specifically, the measure we propose assesses the portion of the lag 12 seasonal autocovariance associated with a given seasonal frequency (spectral peak). Since there are multiple spectral peaks involved, we utilize the same method as McElroy and Holan (2009) to address this multiple testing problem. The issue is that when several independent tests are considered collectively, it is likely that some will grant significant p-values simply due to chance; classically one way this is addressed is by using the Family-Wise Error Rate (FWER) method proposed by Hochberg (1988).

There are other approaches to spectral peak detection in the literature – see the discussion in McElroy and Holan (2009) – but the only ones commonly used in the seasonal adjustment community (leaving aside model-based diagnostics) are VS, the M7 and M8 statistics of X-12 ARIMA, and the tests for stable seasonality of Lytras, Feldpausch, and Bell (2007). These latter tests are not based explicitly on the spectrum, and don't have an ability to discriminate between the various seasonal peaks (i.e., they are all or nothing). We here focus upon procedures that can give detailed information on each seasonal peak: VS, MH, and the diagnostics of this paper. We begin by motivating our statistics through the very intuitive notion of seasonal lag autocovariances, thereby relating the time domain to the spectral density. Section 2 develops this construction into our spectral peak test statistics, and we also discuss a practical solution to the multiple testing problem in this context. Then Section 3 explores the size and power of our tests by evaluating them on the 130 series investigated in McElroy and Holan (2009). Section 4 summarizes our findings: the size and power of the procedure are extremely similar to those of McElroy and Holan (2009) when using the TH kernel, and ignoring the slope test. Therefore, based on interpretation and ease of implementation, we recommend the simplified procedure of this paper for detecting seasonality.

2 Theory

This section discusses the basic methodology for spectral peak testing in this paper. We first discuss the construction of spectral peak test statistics (Section 2.1), and then discuss some theory for the joint testing problem (Section 2.2).

2.1 Spectral Peak Test Statistics

Let X_t be a stationary mean zero time series. (We develop the theory for the null hypothesis – see below – that there are no peaks, or that the spectrum is locally flat; therefore it suffices to consider the time series to be stationary, since a nonstationary time series will not have a flat spectrum. In practice, the method can and will be applied to nonstationary time series.) To focus the discussion, suppose that this is sampled at a monthly frequency, but generalizations to quarterly data will be obvious. We start by observing that the theoretical autocovariance at lag 12 is given by

$$\gamma_{12} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\lambda) \cos(12\lambda)\, d\lambda,$$

where f is the spectral density of the process. It is well-known (Findley et al., 1998) that the spectral density f of a seasonal process will have peaks at some or all of the so-called seasonal frequencies, namely $\pi j/6$ for $j = 1, 2, 3, 4, 5, 6$ (the seasonal frequency at π cannot be identified as a peak due to aliasing, and so will be omitted from further discussion). Due to its oscillating shape, the function $\cos(12\lambda)$ amplifies any of these peaks (as well as the trend frequency at $\lambda = 0$), but dampens the function f in-between the seasonal frequencies. If f has seasonal troughs rather than peaks (a phenomenon not uncommon in seasonally adjusted time series – see McElroy (2008)), these will likewise be amplified to produce a negative value for γ_{12}. Essentially, $\cos(12\lambda)$ acts as a spectral window to extract periodic information from f.

However, because $\cos(12\lambda)$ takes on negative values, it is possible for a spectral peak in one location to be more or less canceled out by a spectral trough at another seasonal frequency. In order to isolate the behavior at each seasonal frequency, we will divide $\cos(12\lambda)$ up into disjoint bands A_1, A_2, A_3, A_4, A_5, such that $A_j = [\pi j/6 - \pi/6, \pi j/6 + \pi/6] \bigcup [-\pi j/6 - \pi/6, -\pi j/6 + \pi/6]$. The union of these five sets covers all of $[-\pi, \pi]$, except for the territory covered by the zero and π frequencies. Then the parameter

$$\rho_j = \frac{1}{2\pi} \int_{A_j} f(\lambda) \cos(12\lambda)\, d\lambda$$

measures the peak/trough of f at frequency $\pi j/6$. We pursue estimation of these parameters nonparametrically, and the simplest estimator is obtained by plugging in the periodogram $I_n(\lambda) = n^{-1} |\sum_{t=1}^{n} e^{-i\lambda t}|^2$, based on a sample of size n. This results in a consistent estimator under very mild conditions (without having the confusion of bandwidth selection that plagues nonparametric spectral density estimation), but we motivate the choice even further. The lag 12 sample autocovariance is simply

$$R_{12} = \frac{1}{\pi} \int_{-\pi}^{\pi} I_n(\lambda) \cos(12\lambda)\, d\lambda,$$

from which it is plain that the estimator

$$\widehat{\rho}_j = \frac{1}{\pi} \int_{A_j} I_n(\lambda) \cos(12\lambda)\, d\lambda$$

is very natural. Furthermore, using the periodogram provides peak/trough estimates that are asymptotically independent, as the following theorem demonstrates.

Theorem 1 *Suppose that the time series is a sample from a causal linear process with fourth moments. Then as $n \to \infty$*

$$\sqrt{n}\, \frac{(\widehat{\rho}_j - \rho_j)}{\sqrt{\frac{2}{\pi} \int_{A_j} I_n^2(\lambda) \cos^2(12\lambda)\, d\lambda}} \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0,1)$$

jointly in $j = 1, 2, 3, 4, 5$, i.e., the statistics are asymptotically independent.

Remark 1 The conditions can be relaxed somewhat: we require that the fourth cumulants of the process vanish and that either condition (B) or (HT) of Taniguchi and Kakizawa (2000, Section 3.1.1) hold, and also that Assumption 1 (8) of Chiu (1988) holds. See the Appendix of McElroy and Holan (2009) and the above papers for more details.

In order to apply this result to peak/trough detection, we must first formulate a null hypothesis. Noting that a flat spectrum on the set A_j produces $\rho_j = 0$, we let the null denote the hypothesis that no peak or trough is present; a large positive (negative) value of ρ_j indicates a peak (trough). Hence for peak detection we have

$$H_0^{(j)} : \rho_j = 0$$
$$H_a^{(j)} : \rho_j > 0$$

for $j = 1, 2, 3, 4, 5$, with the inequality flipped for trough detection. Note that if the spectrum is linear on A_j, then ρ_j will also be zero; it will be nonzero when there is some change in the convexity of f on the interval.

Under each of these null hypotheses, Theorem 1 shows that test statistic defined by

$$T_j = \sqrt{n}\, \frac{\widehat{\rho}_j}{\sqrt{\frac{2}{\pi} \int_{A_j} I_n^2(\lambda) \cos^2(12\lambda)\, d\lambda}}$$

is asymptotically standard normal. Now these statistics are very easy to compute. Let $X = (X_1, X_2, \cdots, X_n)'$ be the data vector, and $R = (R_{1-n}, \cdots, R_0, \cdots, R_{n-1})'$ be the vector of sample autocovariances, where $R_k = n^{-1} \sum_{t=1}^{n-|k|} X_t X_{t+|k|}$. Then the statistic is

$$T_j = n^{-1/2}\, \frac{X' \Gamma(\cos(12\cdot) 1_{A_j}) X}{\sqrt{R' \Gamma(\cos^2(12\cdot) 1_{A_j}) R}},$$

where $\Gamma(g)$ is the Toeplitz matrix associated to g, such that the j,kth entry corresponds to the inverse Fourier Transform of g at index $j-k$. In the cases of g given by $\cos(12\cdot)1_{A_j}$ and $\cos^2(12\cdot)1_{A_j}$ respectively, these matrix entries are just the corresponding autocovariances (given below), and the matrices have dimensions n and $2n-1$ respectively. Direct calculation yields

$$\Gamma_{k,l}(\cos(12\cdot)1_{A_j}) = \begin{cases} \frac{2(k-l)}{\pi}\cos(2\pi j(k-l)/12)\sin(\pi(k-l)/12)/(144-(k-l)^2) & (k-l)\notin 12\mathbf{Z} \\ 1/12 & (k-l)=\pm 12 \\ 0 & \text{else} \end{cases}$$

$$\Gamma_{k,l}(\cos^2(12\cdot)1_{A_j}) = \begin{cases} \frac{(k-l)}{\pi}\cos(2\pi j(k-l)/12)\sin(\pi(k-l)/12)/((k-l)^2-576) & \\ \quad +\sin(\pi(k-l)/12)/(\pi(k-l))\cos(\pi j(k-l)/6) & (k-l)\notin 12\mathbf{Z} \\ 1/12 & (k-l)=0 \\ 1/24 & (k-l)=\pm 24 \\ 0 & \text{else} \end{cases}$$

This enables one to easily encode the entire procedure.

2.2 Joint Peak Testing

We have five peak tests used simultaneously, which can lead to spurious conclusions if the p-values are not interpreted appropriately; this is a well-known multiple testing problem. This issue was addressed in McElroy and Holan (2009) by using the FWER approach of Hochberg (1988) and Benjamini and Hochberg (1995); we use the same method here, detailed as follows. Suppose that we test each of the five peaks individually, as discussed in the previous section, and obtain five p-values. Order these such that $p_{(1)} \leq p_{(2)} \leq p_{(3)} \leq p_{(4)} \leq p_{(5)}$, where $p_{(i)}$ is the ith largest p-value with corresponding null hypothesis $H_{(i)}$ (note the different notation from the previous section, where the superscript on the hypothesis refers to a specific peak, but now the subscript refers to the ordering of the p-values). Then if the FWER is specified at level α (e.g., $\alpha = 0.05$), let k be the largest i for which $p_{(i)} \leq i/(6-i)\,\alpha$. The procedure then indicates that all $H_{(i)}$ are rejected for $i \leq k$.

This Hochberg Family-Wise Error Rate (H-FWER) procedure will make a Type I error, i.e., identify *at least one* seasonal frequency as having a peak when none are actually present, with probability α, so long as the test statistics are independent. Moreover, its statistical power is greater than other methods (Benjamini and Hochberg, 1995). By Theorem 1 the test statistics are asymptotically independent, and hence the size will be approximately correct. The size of the procedure in McElroy and Holan (2009) was inexact due to the two stage testing procedure (i.e., the slope testing aspect).

We also remark here that some practitioners do not regard the fifth seasonal frequency as relevant (Findley, 2006), and hence our peak testing procedure could then be based only on the first four seasonal frequencies. The H-FWER method could then be adapted in an obvious fashion.

3 Empirical Studies

First we consider finite sample size and power for the test statistics T_j using AR(2) processes for the alternative with various frequency locations. This is discussed in 3.1; then in 3.2 we revisit the 130 series of McElroy and Holan (2009), illustrating our peak detection method in practice.

3.1 Size and Power

To evaluate the performance of our diagnostics we conducted several simulations. The first set of simulations examines size (level) for the single peak diagnostic. The results here are straight-forward to obtain, and are similar to the convexity size results for the Tukey-Hanning (TH) kernel of McElroy and Holan (2009). We simulated 10,000 Gaussian white noise series, and evaluated our peak testing method, with results reported in Table 1. While there are some discrepancies for small samples sizes, the convergence is reasonable for $n = 360$.

Next, we consider the empirical power for our single peak diagnostic. As usual, this depends on the chosen α-level for Type I error, as well as on the dynamic range and spectral mass of the underlying peak. We consider $\alpha = .05, .10$ and peaks parametrized by a parameter ϕ as follows: consider the $AR(2)$ process given by

$$(1 - 2\phi \cos \omega B + \phi^2 B^2) X_t = \epsilon_t \tag{1}$$

with white noise variance σ^2, associated with some fixed frequency $\omega \in [0, \pi]$. The spectrum associated with the process in (1) is given by $f(\lambda) = \sigma^2 \left|1 - 2\phi \cos \omega e^{-i\lambda} + \phi^2 e^{-2i\lambda}\right|^{-2}$, which is maximized at $\lambda_0 = \cos^{-1}(\cos \omega (1 + \phi^2)/2\phi)$. Therefore one can explore the power of a peak-testing procedure by simulating from (1) with various choices of ϕ, ω, and σ. Naturally, ω can be set to $\pi j/6$ for $j = 1, 2, 3, 4, 5$ so long as ϕ is fairly close to unity, since then ω will approximately correspond to the peak location. Table 2 presents the result of 10,000 simulations of various sample sizes, from the $AR(2)$ cycle model given in (1) with peak at $\mu = \pi/6$. The peak strength is parametrized through $\phi = .85, .90, .95$. Clearly H_a is true for these processes, and it should be easier to reject H_0 as ϕ is increased towards unity.

As expected, the power of our diagnostic increases with sample size and peakedness, ranging from .145 ($\alpha = .05$) in small samples having a weak spectral peak to perfect power 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, the single peak test possesses decent size and power properties.

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 a process with (stationary) spectral peaks that are realistic, or close to what might be found in practice; in order to have results comparable to McElroy and Holan (2009), we use the same Data Generation Process.

We used $\alpha = .05$, and considered various sample sizes as in our single peak studies. The five single peak test statistics were evaluated on the white noise and $AR(25)$ data respectively, standard normal p-values were calculated, and the H-FWER method discussed in Section 2.2 was applied. If any peaks were detected, this indicated a Type I error for the white noise, but was a correct decision for the $AR(25)$; in this way the empirical size and power were calculated, displayed in Table 3. The size is slowly approaching the nominal level – the discrepancy is essentially due to the lack of independence of the single peak tests in finite sample. The power is exceptional; this is the noted advantage of the H-FWER method. It is also worth contrasting this fairly simple situation with the complexity of McElroy and Holan (2009), where the presence of the slope tests greatly confuses the interpretation of size and power.

3.2 Data

We also considered 130 time series, 65 from the U.S. Census Bureau and 65 from OECD; these are the same data considered in McElroy and Holan (2009), and are repeated here to facilitate comparisons of the two methods. 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. We employed the H-FWER procedure at $\alpha = .05, .10$. For both the raw and seasonally adjusted data, a single trend difference was used (as is the case for the VS diagnostic) before applying the seasonal peaks test.

For Tables 4-9, 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. These tables are simpler than the corresponding ones in McElroy and Holan (2009), which also contain results for VS and other seasonality diagnostics. 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 our method correctly identified seasonality, and hence this proportion is a crude proxy for empirical power. 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.

Our procedure indicates a few cases (when $\alpha = .10$) 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 is at least .69, with higher power for the $\alpha = .10$ level, as expected. These results are quite similar to those of McElroy and Holan (2009) using the TH kernel, although there are a few differences.

4 Conclusion

This paper proposes test statistics that are much simpler to implement than those of McElroy and Holan (2009), and yet serve to identify seasonal peaks with statistical significance. Our approach to the multiple testing problem is the same, but because there is now only one test rather than two, the results are easier to interpret; actual size and power results are similar to those of the convexity test of McElroy and Holan (2009).

References

[1] 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.

[2] Chiu, S. (1988) Weighted least squares estimators on the frequency domain for the parameters of a time series. *The Annalso of Statistics* **16**, 1315–1326.

[3] Findley, D. (2006) Personal communication.

[4] 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).

[5] Hochberg, Y. (1988) A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**, 800–802.

[6] Lytras, D., Feldpausch, R., and Bell, W. (2007) Determining seasonality: a comparison of diagnostics from X-12-ARIMA. 2007 Proceedings of the Third International Conference on Establishment Surveys. http://www.census.gov/srd/www/sapaper/sapaper.html

[7] McElroy, T. (2008) A modified model-based seasonal adjustment that reduces spectral troughs and negative seasonal correlation. *SRD Research Report No.* $RRS2008-12$, U.S. Census Bureau.

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

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

Size - Single Peak Test						
	$\alpha = .05$			$\alpha = .10$		
n	Size	mean	st. dev.	Size	mean	st. dev.
120	.014	-.011	.892	.075	-.010	.884
144	.020	-.022	.895	.076	-.022	.889
180	.032	-.050	.929	.082	-.010	.903
288	.037	-.021	.940	.088	-.022	.942
360	.033	-.060	.961	.092	-.005	.942

Table 1: Simulated size results for $\alpha = .05$ and $\alpha = .10$ based on 10,000 repetitions of Gaussian white noise, at various sample sizes.

Power - Single Peak Test						
	$\alpha = .05$			$\alpha = .10$		
n	$\phi = .85$	$\phi = .90$	$\phi = .95$	$\phi = .85$	$\phi = .90$	$\phi = .95$
120	.145	.319	.701	.363	.614	.905
144	.188	.457	.799	.410	.698	.950
180	.286	.586	.911	.489	.783	.981
288	.472	.824	.995	.652	.916	.999
360	.551	.907	1.00	.740	.961	1.00

Table 2: Simulated power results for $\alpha = .05$ and $\alpha = .10$ based on 10,000 repetitions of a Gaussian $AR(2)$ process, at various sample sizes and peak strengths.

Size and Power - Joint Test				
	$\alpha = .05$		$\alpha = .10$	
n	Size	Power	Size	Power
120	.002	.892	.048	.997
144	.003	.952	.051	1.00
180	.005	.991	.062	1.00
288	.015	1.00	.083	1.00
360	.021	1.00	.086	1.00

Table 3: Simulated size and power results for joint peak testing, with $\alpha = .05$ and $\alpha = .10$ based on 10,000 repetitions of Gaussian white noise and a multi-peak $AR(25)$ process, at various sample sizes. The H-FWER procedure (Section 2) was used to determine rejection of the null hypotheses.

Manufacturing Series				
	$\alpha = .05$		$\alpha = .10$	
series	**Original Data**	**SA Data**	**Original Data**	**SA Data**
M_1	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_2	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_3	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_4	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
M_5	1,2,3,4,5	$\emptyset$	1,2,3,4,5	1,2,3,4,5
M_6	$\emptyset$	$\emptyset$	1,2,3	$\emptyset$
M_7	$\emptyset$	$\emptyset$	1,2,4	$\emptyset$
M_8	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_9	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{10}	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
M_{11}	$\emptyset$	$\emptyset$	1,2,3,5	$\emptyset$
M_{12}	1,2,3,4	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{14}	$\emptyset$	$\emptyset$	1,2,3,4	$\emptyset$
M_{15}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{16}	1,2,3,4	$\emptyset$	1,2,3,4	$\emptyset$
M_{17}	$\emptyset$	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{18}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{19}	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
M_{20}	$\emptyset$	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{21}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{22}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{23}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{24}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{25}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{26}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{27}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{28}	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
M_{29}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{30}	2,3,4,5	$\emptyset$	1,2,3,4,5	1,3,4,5
M_{31}	$\emptyset$	$\emptyset$	1,2,3	$\emptyset$
M_{32}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{33}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{34}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{35}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M_{36}	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
Total "correct"	24/35	35/35	31/35	33/35

Table 4: Data analyses for 35 Manufacturing Series (U.S. Census Bureau) comparing our multiple peak diagnostic. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = .05$ and $\alpha = .10$.

Housing Series				
	$\alpha = .05$		$\alpha = .10$	
series	**Original Data**	**SA Data**	**Original Data**	**SA Data**
MW1Fam	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
MWTot	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
NE1Fam	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
NETot	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
S1Fam	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
STot	1,2,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
US1Fam	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
USTot	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
W1Fam	1,2,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
WTot	1,2,4,5	$\emptyset$	1,2,4,5	$\emptyset$
Total “correct”	10/10	10/10	10/10	10/10
Import/Export Series				
	$\alpha = .05$		$\alpha = .10$	
series	**Original Data**	**SA Data**	**Original Data**	**SA Data**
M00120	$\emptyset$	$\emptyset$	1,2,3,4,5	$\emptyset$
M00190	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M3000	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M3010	2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
M12060	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
X3	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
X00300	2,3,4,5	$\emptyset$	2,3,4,5	$\emptyset$
X3020	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
X3022	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
X10140	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
Total “correct”	9/10	10/10	10/10	10/10
Retail Series				
	$\alpha = .05$		$\alpha = .10$	
series	**Original Data**	**SA Data**	**Original Data**	**SA Data**
s0b441x0	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
s0b44000	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
s0b44100	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
s0b44130	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
s0b44200	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
s0b44300	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
s0b44312	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
s0b44400	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
s0b44410	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
s0b44500	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
Total “correct”	10/10	10/10	10/10	10/10
Grand Total “correct”	29/30	30/30	30/30	30/30

Table 5: Data analyses for 30 U.S Census Bureau Series (10 Housing, 10 Import/Export and 10 Retail Sales) comparing our multiple peak diagnostic. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = .05$ and $\alpha = .10$.

OECD DEU				
	$\alpha = .05$		$\alpha = .10$	
series	**Original Data**	**SA Data**	**Original Data**	**SA Data**
PRMNCG03	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNCS01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNIG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNVG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLMNCD01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLMNCN01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLMNDM01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLMNEX01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	1,2,3,4,5
SLMNIG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLMNTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLRTCR01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	4,5
SLRTTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLRTTO02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLWHTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	2,3,4
SLWHTO02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	2,3,4
UNLVRG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
UNLVSUMA	1,2,3,4,5	$\emptyset$	1,2,3,4,5	1,2,3,4,5
UNLVSUTT	1,2,3,4,5	$\emptyset$	1,2,3,4,5	1,2,3,4,5
UNRTRG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	1,2,3,4,5
XTEXVA01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
XTIMVA01	2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
Total "correct"	22/22	22/22	22/22	15/22

Table 6: Data analyses for 22 German OECD Series comparing our multiple peak diagnostic. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = .05$ and $\alpha = .10$.

OECD EMU				
	$\alpha = .05$		$\alpha = .10$	
series	**Original Data**	**SA Data**	**Original Data**	**SA Data**
PRCNTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRINTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNCG03	1,2,3,4,5	4	1,2,3,4,5	4
PRMNCS01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNIG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNVG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLMNCN02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLMNIG02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	1,2,3,4,5
SLMNTO02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	2,3,4,5
SLMNVG02	1,2,3,4,5	2	1,2,3,4,5	1,2,3,4,5
SLRTTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLRTTO02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
XTEXVA01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	2,4,5
XTIMVA01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
Total "correct"	15/15	13 /15	15/15	10/15

Table 7: Data analyses for 15 Euro-area OECD Series comparing our multiple peak diagnostic. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = .05$ and $\alpha = .10$.

OECD FRA				
	$\alpha = .05$		$\alpha = .10$	
series	**Original Data**	**SA Data**	**Original Data**	**SA Data**
PRAFAG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRCNTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNCG01	1,2,3,4,5	4	1,2,3,4,5	4
PRMNCS01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNIG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNTO01	1,2,3,4,5	3,4,5	1,2,3,4,5	1,2,3,4,5
PRMNVE01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	1,2,3,4,5
SLRTCR01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLRTTO02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
XTEXVA01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
XTIMVA01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
Total "correct"	11/11	9 /11	11/11	8/11

Table 8: Data analyses for 11 French OECD Series comparing our multiple peak diagnostic. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = .05$ and $\alpha = .10$.

OECD GBR				
	$\alpha = .05$		$\alpha = .10$	
series	**Original Data**	**SA Data**	**Original Data**	**SA Data**
PPIAMP01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PPIAMP02	3	$\emptyset$	1,2,3,4,5	$\emptyset$
PPIPFU01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRINTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNCG02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNCG03	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNCS01	2,4	$\emptyset$	2,3,4,5	$\emptyset$
PRMNIG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNTO01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNVE02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNVE03	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
PRMNVG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
SLRTCR03	1,2,3,4,5	2,3,4,5	1,2,3,4,5	1,2,3,4,5
SLRTTO02	1,2,3,4,5	$\emptyset$	1,2,3,4,5	1,2,3,4,5
UNLVRG01	1,2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
XTEXVA01	2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
XTIMVA01	2,3,4,5	$\emptyset$	1,2,3,4,5	$\emptyset$
Total "correct"	17/17	16/17	17/17	15/17

Table 9: Data analyses for 17 Great Britain OECD Series comparing our multiple peak diagnostic. Our multiple peak diagnostic uses the H-FWER method to control the FWER at $\alpha = .05$ and $\alpha = .10$.