RESEARCH REPORT SERIES
*(Statistics #2017-03)*

# Detecting Seasonality in Seasonally Adjusted Monthly Time Series

David F. Findley [1]
Demetra P. Lytras
Tucker S. McElroy

[1] Retired from the U.S. Census Bureau

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

Report Issued: February 13, 2017

# Detecting Seasonality in Seasonally Adjusted Monthly Time Series

David F. Findley, Demetra P. Lytras, Tucker S. McElroy
U.S. Census Bureau (Retired), U.S. Census Bureau, U.S Census Bureau
David.Findley@ieee.org, Demetra.P.Lytras@census.gov, Tucker S. McElroy@census.gov

February 1, 2017

**Abstract**

The most fundamental seasonal adjustment deficiency is detectable seasonality after adjustment. Residual seasonality has reduced amplitudes and other properties which make it necessary to undertake its detection differently from seasonality detection in unadjusted series. We present the results of our investigation of residual seasonality detection properties of three types of diagnostics, regression, spectrum and positive seasonal autocorrelation, six diagnostics in all. These are available, sometimes with modifications, in widely used software, five in TRAMO-SEATS, X-13ARIMA-SEATS and JDemetra+, one only in JDemetra+. All were applied to a set of sixteen underadjusted U.S. Census Bureau Monthly Retail Trade Survey series. The series have evidence of changing seasonality but were deliberately adjusted only for stable seasonality. Residual seasonality was found in the final 8 years of all series and all but one irregulars series. Patterns of "weak" and "strong" detections differ by diagnostic and final span length, sometimes in quite complementary ways.

**Keywords** Seasonality diagnostics · Residual seasonality



## Contents

**1 Introduction and Overview** **2**

**2 The Spectral Density and Related Diagnostics** **5**
2.1 The Autoregressive Spectrum Diagnostic . . . . . . . . . . . . . . . . . . . . 7
2.1.1 The Visual Significance Criterion . . . . . . . . . . . . . . . . . . . . 7
2.1.2 AR Spectrum Detection Background . . . . . . . . . . . . . . . . . 8
2.2 Sample Spectral Density and Periodogram Based Diagnostics . . . . . . 9
2.2.1 The Periodogram Sum Diagnostic . . . . . . . . . . . . . . . . . . 10
2.3 Consistent Spectrum Estimators and the $M$-$T$ Diagnostic . . . . . . . . 10
2.4 A First Example of Residual Seasonality Detection . . . . . . . . . . . . 11

**3 The RegARIMA GLS F-Statistic for Stable Seasonality** **16**
3.1 What do Stable-Seasonal Regression Coefficients Estimate? . . . . . . . 18

**4 Maravall's QS Statistic for Positive Seasonal Autocorrelation** **21**
4.1 Background for QS with Residual Seasonality . . . . . . . . . . . . . . . . 22

**5 Empirical Residual Seasonality Detections** **23**
5.1 Detections from the Stable-Seasonal Adjustments . . . . . . . . . . . . . 25
5.2 Detections from the Associated Irregulars . . . . . . . . . . . . . . . . . . 28

**6 Concluding Remarks** **30**

**7 Appendix** **31**
7.1 Inconsistency and Other Properties of $\tilde{g}(\lambda)$ . . . . . . . . . . . . . . . . . 31
7.2 The Calendar-Month-Average Stable Seasonal Component . . . . . . . . 32
7.3 Insights from SAR(1) Series . . . . . . . . . . . . . . . . . . . . . . . . . . 33
7.3.1 Analysis of the SAR(1) Stable-Seasonal Component . . . . . . . 34

## 1. Introduction and Overview

The most fundamental seasonal adjustment deficiency is detectable residual seasonality in the adjusted series. With direct additive or log-additive X-11 or ARIMA model-based seasonal adjustments (and approximately with multiplicative X-11 adjustments), Bell (2011) shows that the seasonal adjustment filters of all widely used seasonal adjustment software remove stable (fixed) seasonality. As we illustrate, this includes removal of the stable seasonality defined by each calendar month's average seasonal effect over the data span adjusted. These results apply to the seasonal adjustments of X-13ARIMA-SEATS (U.S. Census Bureau, 2016), hereafter X-13A-S, TRAMO-SEATS (Gómez and Maravall, 1996) and TSW (Caporello and Maravall, 2004), hereafter T-S, and JDemetra+ (National Bank of Belgium, 2015), hereafter JD+. Consequently, with this software, except perhaps with series that have very volatile calendar month subseries, residual seasonality can only occur when seasonality is changing over time, a property of all residual seasonality.

There are several possible causes of residual seasonality, starting with inadequate option or model specification. It can also happen that the series has level or calendar month movements so erratic or substantial over the length of data specified for adjustment that no software options, or no fixed-coefficient ARIMA model with an admissible seasonal decomposition, can produce seasonal adjustment filters that are effective for the full data span specified for adjustment.

The absence of a stable-seasonal component causes residual seasonality to be generally weaker and more difficult to detect than seasonality in an unadjusted series. This is why its detection requires a separate study like that presented here. The same basic diagnostics can be used with little or no modification. But for best residual seasonality detection, our results show that most diagnostics should be calculated for an appropriate subspan of the adjusted series (or of its irregulars), see the graphs and discussion of Subsection 3.1.

Apart from the new JD+ periodogram sum OLS F-test diagnostic of De Antonio and Palate (2014), all seasonal adjustment programs mentioned have the diagnostics we consider (or slight variants thereof). Some have more.

The frequency domain diagnostics (spectrum, periodogram, periodogram sum) are described in Section 2, the regression diagnostics in Section 3, and Maravall's QS positive seasonal autocorrelation diagnostic in Section 4.

Table 1 lists the diagnostics by type, with an indication of each diagnostic's seasonality detection criterion. The degrees of freedom formulas are somewhat complicated and not fully explained until later. They can be ignored until such details are of interest to the reader. The "visual significance" criterion is defined in Subsection 2.1 and illustrated in the figures of Subsection 2.4.

For a specified interval of seasonally adjusted or irregular series values of length $n$ months, with $n$ large enough for diagnostic estimation ($n \geq 96$ preferred), the diagnostics are calculated from the *stationarized* values, obtained by applying differencing $(1-B)^d$ and log transformation as appropriate to the $n-d$ most recent values of the specified interval. ($d=0$ for irregulars.) This yields an interval of values for diagnostic calculation of length $n-d$, for the periodogram sum diagnostic length $n-d+h, -12<h<12$ with $n-d+h$ divisible by 12, as explained in Subsection 2.2.1. The values of $d$ and $h$ are diagnostic-dependent. For seasonally adjusted values, $d$ is related to the seasonal and nonseasonal differencing orders of the ARIMA model for the data, see U.S. Census Bureau (2016, Chap. 7.17). For $h$ see Subsection 2.2.

The JD+ periodogram sum diagnostic is equivalent to an Ordinary Least Squares (OLS) regression F-test diagnostic for the interval of length $n-d+h$. It is denoted by $F^{fs}$. Table 1 shows two other diagnostics also associated with $F$ statistics. More details are given after the table and in the sections presenting diagnostics grouped by type.

**Table 1. Diagnostics for a Specified SA or Irregulars Interval of Length $n$**

| Seasonality Diagnostic | Seasonality Criterion |
|---|---|
| AR(30) spectrum seas. freq. peaks | "Visual Significance" |
| Periodogram seas. freq. peaks | "Visual Significance" |
| Maravall "Tukey" spectrum ratios | Rejection of Quasi-$F_{df1,df2}$ ($df1, df2$ from table-fit functions) |
| Periodogram sum F-test | Rejection of OLS $F^{fs}_{df1,df2}$ ($df1 = 11 - 1^{\text{even}}_{n-d+h}$, $df2 = n-d+h-12+1_{n-d-h \text{ even}}$) |
| Stable-seasonal regression | Rejection of GLS $F^{M}_{df1,df2}$ ($df1 = 11, df2 = n-d-k$) |
| Positive seas. autocorrelation $QS$ | Rejection of Quasi-$\chi^2(2)$ |

The first three diagnostics evaluate seasonal frequency amplitudes relative to neighboring frequency amplitudes using a ratio criterion, empirical or simulation-based.

In the degrees of freedom formulas for the periodogram sum diagnostic, the indicator $1^{\text{even}}_{n-d+h}$ has the value 1 if $n-d+h$ is even and 0 otherwise. For the stable-seasonal regression diagnostic, $k \geq 11$ is the total number of independent regressors in the regARIMA model estimated, including any holiday, trading day and outlier regressors in addition to the eleven seasonal regressors (14).

Section 4 concerns the $QS$ diagnostic of Maravall (2012) for detecting positive seasonal autocorrelation after appropriate non-seasonal differencing. The qualifier "quasi-" is used in Table 1 when, as with $QS$, the null hypothesis distribution indicated is a simulation-based approximation to the actual distribution.

Section 5 presents the diagnostics' performance results for a set of sixteen *deliberately underadjusted* historical U.S. Retail Trade Survey[1] series. The series start in January 1992 and end in December 2007, prior to the Great Recession. They have moving seasonality over this interval but were adjusted only for stable seasonality. The diagnostics are applied to different-length subspans of the stable-seasonal adjustment of the full data span of each series as well as to the corresponding subspans of each adjustment's irregulars. The subspans we focus on have January starting months and final month December 2007. Throughout, $n$ denotes the length of the subspan specified for seasonality detection. For $n = 96$, with starting date January 1990, results for all diagnostics from all series are given in Tables 4 and 6. For other starting dates, only each diagnostic's weak and strong (see Table 3 of Section 5) detection totals for each $n$ are given, in Tables 5 and 7, not the individual series results.

A reader with some exposure to spectrum diagnostics and to regression estimation of seasonality could start from the expository examples in Subsections 2.4 and 3.1. Then, after examining Subsection 4.1 about why positive first-seasonal-lag autocorrelation is especially problematic in seasonally adjusted series (appropriately differenced) and irregulars, the diagnostics' detection results for the sixteen series could be examined, consulting earlier subsections for more details as needed.

## 2. The Spectral Density and Related Diagnostics

Because seasonality induces quasi-repetitive year to year movements in a time series, frequency domain diagnostics are natural for stationary series, including stationarized series resulting from differencings and log transformation as appropriate. These diagnostics express time series properties in units of cycles per year. Estimators of the spectral density, which does this for the autocovariance properties of stationary

[1] This is a sample survey subject to both sampling and nonsampling error. A description of the survey methods is available at https://www.census.gov/retail/mrts/how_surveys_are_collected.html.

series, are traditional choices. For a stationary series $x_t$ with mean $\mu = Ex_t$, set $\gamma_k = E(x_t - \mu)(x_{t-k} - \mu), k = 0, \pm 1, \pm 2, \ldots$ .

The spectral density function $g(\lambda)$, abbreviated *sd*, is defined for $-1/2 \leq \lambda \leq 1/2$ by

$$g(\lambda) = \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \left( e^{i2\pi k\lambda} + e^{-i2\pi k\lambda} \right) = \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos 2\pi k\lambda. \quad (1)$$

A spectral density is an even function, $g(-\lambda) = g(\lambda)$ for all $\lambda$. Consequently, for graphs, peak heights, and the other features of interest for our study, we can can restrict attention to $0 \leq \lambda \leq 1/2$. Each autocovariance $\gamma_k, k = 0, \pm 1, \pm 2, \ldots$ can be recovered from the spectral density via

$$\gamma_k = \int_{-1/2}^{1/2} e^{i2\pi k\lambda} g(\lambda)\, d\lambda = 2 \int_0^{1/2} \cos 2\pi k\lambda\, g(\lambda)\, d\lambda, \quad (2)$$

see Brockwell and Davis (1991, 120). We assume $\mu = 0$. An uncorrelated stationary series with mean zero is called *white noise*. It follows from (1) and (2) that, among stationary zero-mean series, only white noise has a constant spectral density, $g(\lambda) = \gamma_0$, the variance.

With a monthly series, $\lambda = 1/12$ cycles/month is the *fundamental seasonal frequency*, corresponding to one cycle every 12 months. The remaining *seasonal frequencies*, $\lambda_j = j/12,\ 2 \leq j \leq 6$, often called harmonics, are the frequencies with an integral number cycles in 12 months.

For an ARMA process $\phi(B)(x_t) = \theta(B)\, a_t$ with $var(a_t) = \sigma_a^2$,

$$g(\lambda) = \sigma_a^2 \frac{\left|\theta\left(e^{i2\pi\lambda}\right)\right|^2}{\left|\phi\left(e^{i2\pi\lambda}\right)\right|^2}, \quad 0 \leq \lambda \leq 1/2, \quad (3)$$

see[2] Brockwell and Davis (1991, 123). Thus, for a monthly SAR(1) process,

$$x_t = \Phi x_{t-12} + a_t, |\Phi| < 1, \quad (4)$$

$$g(\lambda) = \sigma_a^2 \left|1 - \Phi e^{i2\pi 12\lambda}\right|^{-2} = \left(1 + \Phi^2 - 2\cos 2\pi 12\lambda\right)^{-2}.$$

[2] For the reader having little familiarity with complex numbers, we note that the magnitude $|z|$ of a complex number $z = x + iy$ is $|z| = \sqrt{x^2 + y^2}$ and, for any real $\lambda$, $e^{i2\pi\lambda} = \cos 2\pi\lambda + i \sin 2\pi\lambda$. See Wikipedia Contributors (2012) for more background. Thus $\left|\theta\left(e^{i2\pi\lambda}\right)\right|^2 = \left|1 - \sum_{j=1}^{r} \theta_j e^{i2\pi j\lambda}\right|^2$ is reexpressible without complex numbers as $\left(1 - \sum_{j=1}^{r} \theta_j \cos 2\pi j\lambda\right)^2 + \left(\sum_{j=1}^{r} \theta_j \sin 2\pi j\lambda\right)^2$.

The SAR(1) has variance $\gamma_0 = Ex_t^2 = \sigma_a^2 \left(1 - \Phi^2\right)^{-2}$ and nonzero autocorrelations only at seasonal lags,

$$\rho_{12j} = \Phi^j, j \geq 1; \quad \rho_l = 0, l \neq 12j. \tag{5}$$

### 2.1. The Autoregressive Spectrum Diagnostic

For monthly data, following Akaike and Ishiguro (1980), in X-13A-S an AR($p$) spectral density estimate is used to detect both trading day effects and seasonal effects, the frequencies for the latter being the seasonal frequencies, $\lambda_k = k/12$, $k$ cycles per year, for $1 \leq k \leq 6$. In order to obtain adequate resolution of all frequencies of interest, an AR($p$) model $\phi\left(B\right) x_t = a_t$ with $p = 30$ is estimated for the stationarized data $x_t$ from the interval being investigated. The Yule-Walker estimation method is used to obtain an estimated AR polynomial $\hat{\phi}\left(B\right) = 1 - \sum_{j=1}^{p} \hat{\phi}_j B^j$ with the required property that $\hat{\phi}\left(z\right) \neq 0$ for $|z| \leq 1$, see §8.1 of Brockwell and Davis (1991). With $\hat{\sigma}_a^2$ denoting the associated estimate of $\sigma_a^2$, from (3), the *sd* estimate is

$$\hat{g}\left(\lambda\right) = \frac{\hat{\sigma}_a^2}{\left|\hat{\phi}\left(e^{i2\pi\lambda}\right)\right|^2}. \tag{6}$$

The properties of an *sd* estimate $\hat{g}\left(\lambda\right)$ are more usefully expressed after a log transform, most commonly the *decibel transform* $10\log_{10}\hat{g}\left(\lambda\right)$. Transformation of (6) to decibel units yields the first of the spectrum diagnostics we consider,

$$arspec\left(\lambda\right) = 10\log_{10}\hat{g}\left(\lambda\right), 0 \leq \lambda \leq 1/2. \tag{7}$$

The seasonal peak significance criterion for *arspec* described next is used for (7) by X-13A-S, T-S, and JD+.

#### 2.1.1. The Visual Significance Criterion

For series with prominent seasonal features, the *sd* estimates we consider, calculated after appropriate differencing, usually show local maxima at one or more seasonal frequencies. There are no well established tests for the statistical significance of such seasonal peaks. Instead, empirically validated criteria are used, such as the one now presented.

For *arspec* the criterion was motivated by the AR spectrum printer plots produced by BAYSEA (Akaike and Ishiguro, 1980). It uses the following definitions of a seasonal peak. At $\lambda_k = k/12 = 10k/120, 1 \leq k \leq 5$, a *seasonal peak* is a value $arspec\,(\lambda_k)$ larger than the neighboring values $arspec\,(\lambda_k \pm 1/120)$. At $\lambda_6 = 1/2$, $arspec\,(\lambda_6) > arspec\,(\lambda_6 - 1/120)$ is required. For $k \neq 6$, a seasonal peak value $arspec\,(\lambda_k)$ is said to be *visually significant*, denoted v.s., if it is larger than $arspec\,(\lambda_k \pm 1/120)$ by at least 6/52 of the range,

$$\max_{0 \leq j \leq 60} arspec\,(j/120) - \min_{0 \leq j \leq 60} arspec\,(j/120)\,,$$

and also larger than the median of $arspec\,(0)\,, arspec\,(1/120)\,, \ldots, arspec\,(60/120)$. (Six *arspec* values at trading day frequencies, which do not have the form $j/120$, replace the $arspec\,(j/120)$ values in the range and median calculations for v.s. trading day effects.) In the *arspec* graphs of Figures 2, 4, and 5, the scale bars at the right show the peak height required for v.s.

This criterion was first used by Soukup and Findley (1999) to detect peaks at frequencies indicative of trading day effects. A v.s. criterion for $\lambda_6 = 1/2$ has not been determined; see Subsection 7.1. (With X-13A-S, other multipliers of 1/52 can be specified, e.g. a 12* v.s. detection criterion.) The v.s. criterion is numerical and does not require graphs. But spectrum graphs can be informative, see Subsection 2.4, and are produced by the programs. For monthly data, the X-13A-S default is to calculate *arspec* from the last 96 values under investigation when this many values are available, and similarly for *pdg* of (12) when it is specified. Results for the last 120 and 144 months are also shown for our empirical study series. It will be seen that *arspec* is competitive with the other spectrum diagnostics considered, whereas *pdg* is not.

### 2.1.2. AR Spectrum Detection Background

Autoregressive spectrum estimation has a tradition of use in electrical engineering and geophysics as an exploratory diagnostic for detecting the presence of periodic components, see Marple (1987). For statistical perspectives, see Berk (1974), Priestley (1981, 600-611), and Newton and Pagano (1983). With a monthly time series, an additive stable-seasonal component $S_t$ of $x_t$ is conceived as being perfectly repetitive with average zero over any 12 month period. Equivalently $S_t = -S_{t-1} - \cdots - S_{t-11}$, an autoregressive relation with no white noise. As our empirical study will show, even when seasonal movements are not perfectly repetitive, a strong tendency for same calendar month values to move in the same direction from one year to the next, a loose kind of positive seasonal correlation, can lead to small values of $\left|\hat{\phi}\left(e^{i2\pi\lambda}\right)\right|^2$ in (6) at one or more seasonal frequencies and therefore to large values of $arspec\left(\lambda\right)$ at these $\lambda$.

### 2.2. Sample Spectral Density and Periodogram Based Diagnostics

The *sample spectral density* $\tilde{g}\left(\lambda\right)$ of $x_1, \ldots, x_N$ (e.g., $N = n - d$) is defined by

$$\tilde{g}\left(\lambda\right) = c_0 + 2\sum_{k=1}^{N-1} c_k \cos 2\pi k\lambda, 0 \leq \lambda \leq 1/2, \tag{8}$$

where

$$c_k = N^{-1}\sum_{t=1}^{N-k}\left(x_t - \bar{x}\right)\left(x_{t+k} - \bar{x}\right), 0 \leq k \leq N-1, \tag{9}$$

with $\bar{x} = N^{-1}\sum_{t=1}^{N} x_t$. A related alternative with a long history has the form

$$I_N\left(\lambda\right) = \frac{1}{N}\left|\sum_{t=1}^{N} x_t e^{-i2\pi t\lambda}\right|^2, 0 \leq \lambda \leq 1/2. \tag{10}$$

For integer $j$, frequencies $\lambda_j = j/N$ in $-1/2 \leq \lambda \leq 1/2$ (in $-N/2 \leq j \leq N/2$ when $N$ is even) are called *Fourier frequencies*. The values $I_N\left(\lambda_j\right)$ define the classical *periodogram*. Proposition 10.1.2 of Brockwell and Davis (1991) shows that

$$I_N\left(\lambda_j\right) = \begin{cases} \hat{g}\left(\lambda_j\right) & , j \neq 0 \\ N\bar{x}^2 & , j = 0 \end{cases}. \tag{11}$$

The function $I_N\left(\lambda\right), -1/2 \leq \lambda \leq 1/2$ of (10) can be called the *all frequency periodogram*. Transformed to decibel units, it defines the X-13A-S periodogram diagnostic provided in response to user requests,

$$pdg\left(\lambda\right) = 10\log_{10} I_N\left(\lambda\right), 0 \leq \lambda \leq 1/2. \tag{12}$$

*pdg* is evaluated at the same frequencies as *arspec*, again using the v.s. criterion. Only under assumptions too restrictive for our application are well-documented hypothesis test statistics for detecting frequencies of individual periodic components in correlated series $x_t$ available for this *pdg*, see Priestley (1981).

### 2.2.1. The Periodogram Sum Diagnostic

Special properties of the classical periodogram are used to obtain the *periodogram sum* OLS F-test diagnostic of De Antonio and Palate (2014), which we denote by $F^{fs}$. Results for $F^{fs}$ are shown in Section 5 along with results for the other spectrum diagnostics. $F^{fs}$ is available only in JD+. It is based on the sum of the values at the six seasonal frequencies of the classical periodogram of an interval of $(1-B)^d, d \geq 0$ differenced data values, log transformed if needed. The most recent value is the same as that of the interval of length $n-d$ used for the other spectrum diagnostics, but $-12 < h < 12$ is chosen so that $n-d+h$ is divisible by 12, in order to have seasonal frequencies be Fourier frequencies, which is seldom true of $n-d$, i.e. for $h=0$. Thus the interval tested by $F^{fs}$ is usually an extension or contraction of the interval being tested by the other frequency diagnostics. If enough preceding time series values are available to calculate the required differenced data, the smallest $0 < h < 12$ providing an extended interval whose length $n-d+h$ is divisible by 12 is used. Otherwise, not all of the available differenced data are used. The subinterval of length $n-d+h$, $-12 < h < 0$, with smallest magnitude $h$ providing divisibility of $n-d+h$ by 12, is used. Because $F^{fs}$ is determined by OLS instead of GLS estimation, no ARMA model is needed for the values of this interval.

The degrees of freedom formulas for $F^{fs}$ are shown in Table 1. We use $F_\alpha^{fs}$ to indicate the test at level $0 < \alpha < 1$. J. Palate (Personal communication, April 5, 2016) provided the $F_\alpha^{fs}$ technical information and results of Section 5. Formal documentation will be available later from https://github.com/jdemetra/jdemetra-core/wiki/Seasonality-tests.

## 2.3. Consistent Spectrum Estimators and the *M*-*T* Diagnostic

Neither $I_n(\lambda)$ nor $\tilde{g}(\lambda)$ converges statistically to $g(\lambda)$ as $n \to \infty$: they are not consistent estimators of $g(\lambda)$. Section 7.1 references interesting properties of the ratio $\tilde{g}(\lambda) / g(\lambda)$ that reveal this. The estimator $\tilde{g}(\lambda)$ can be modified to have a limiting distribution for hypothesis tests. A standard modification applies a weighting function $w(k/M)$ to the $c_k$ in (8), one that gives zero weight to lags $k > m$, with $m = m(n) \to \infty$ chosen so that $m(n)/n^{1/2} \to 0$ as $n \to \infty$. The Blackman-Tukey Hanning estimator applied in the spectrum diagnostic of Maravall (2012) uses $w(v) = 0.5 + 0.5\cos \pi v$ for $0 \leq v \leq 1$

and $w(v) = 0$ for $v > 1$ resulting in

$$h(\lambda) = c_0 + 2\sum_{k=1}^{m} w(k/m)\, c_k \cos 2\pi k\lambda. \tag{13}$$

Under Gaussian ARMA and more general assumptions, for each $\lambda$, $\left(n^{1/2}/m(n)\right)[h(\lambda) - g(\lambda)]$ has a mean zero Gaussian limiting distribution with variance $2g(\lambda)^2 \int_0^1 w^2(\nu)\, d\nu = 0.75 g(\lambda)^2$ for $0 < \lambda < 1/2$, doubling in value at $\lambda = 0, 1/2$, see Theorem 9.4.1 of Anderson (1971).

With $\lambda_j = j/m$, Maravall's Tukey spectrum diagnostic focuses on the statistical significance of

$$H(\lambda_j) = \frac{2h(\lambda_j)}{h(\lambda_{j-1}) + h(\lambda_{j+1})}, 1 \leq j \leq m-1$$

and

$$H(1/2) = h(\lambda_m)/h(\lambda_{m-1}), \qquad j = m,$$

for values of $m$ determined by the interval length $n$. For a monthly series of length $80 \leq n \leq 119$, $m = 79$. For $n \geq 120$, $m = 112$. The choice $m = 79$ is also identified in Maravall (2012) as "the lowest frequency that permits the isolation of trading day peaks from seasonal peaks." We call the resulting estimator the *M-T spectrum diagnostic*, abbreviated *M-T*.

In T-S and X-13A-S, the hypothesis of no significant seasonal peaks is tested at the .01 level using an approximating $F_{df1,df2}$ distribution with $df1$ and $df2$ values for sample sizes $80 \leq n \leq 300$ from functions fit to the sample means and variances of $H(\lambda_j)$ from white noise simulations, see Maravall (2012). Empirical study series results for the *M-T* diagnostic are included in the tables of Section 5.

### 2.4. A First Example of Residual Seasonality Detection

Detections by *arspec* and *pdg* are the most easily visualized: We consider two X-13A-S seasonal adjustments of the 16 year span January 1992–December 2007 of Sales of U.S. Warehouse Clubs and Superstores. The first has residual seasonality. The second, a default automatic adjustment, does not. (All seasonal adjustments considered in this article differ from published Census Bureau seasonal adjustments.) The first adjusts only for stable seasonality, estimated via the X-11 specification `seasonalma=stable`. Its last 8 years are shown in Figure 1 along with the unadjusted series values. Its *arspec* for these 8 years, displayed in Figure 2, has v.s. peaks at the seasonal frequencies 1/12, 2/12 and 5/12 cycles per month. Its *pdg* has only a 2/12 v.s. peak, see Figure 3.

The automatic adjustment of the 16 year span estimates moving seasonality: The moving seasonality ratio filter selection procedure (see Dagum (1980) or Ladiray and Quenneville (2001)), which is specified in X-13A-S by x11{`seasonalma=msr`}, chose the $3 \times 5$ seasonal filter. For the last 8 years of this adjustment, neither *arspec* nor *pdg* has a v.s. peak. We only show *arspec*, see Figure 5. With Figure 4 as a reference, the residual seasonality of the adjusted series in Figure 1 is visible as substantial downward movements around every December or January.

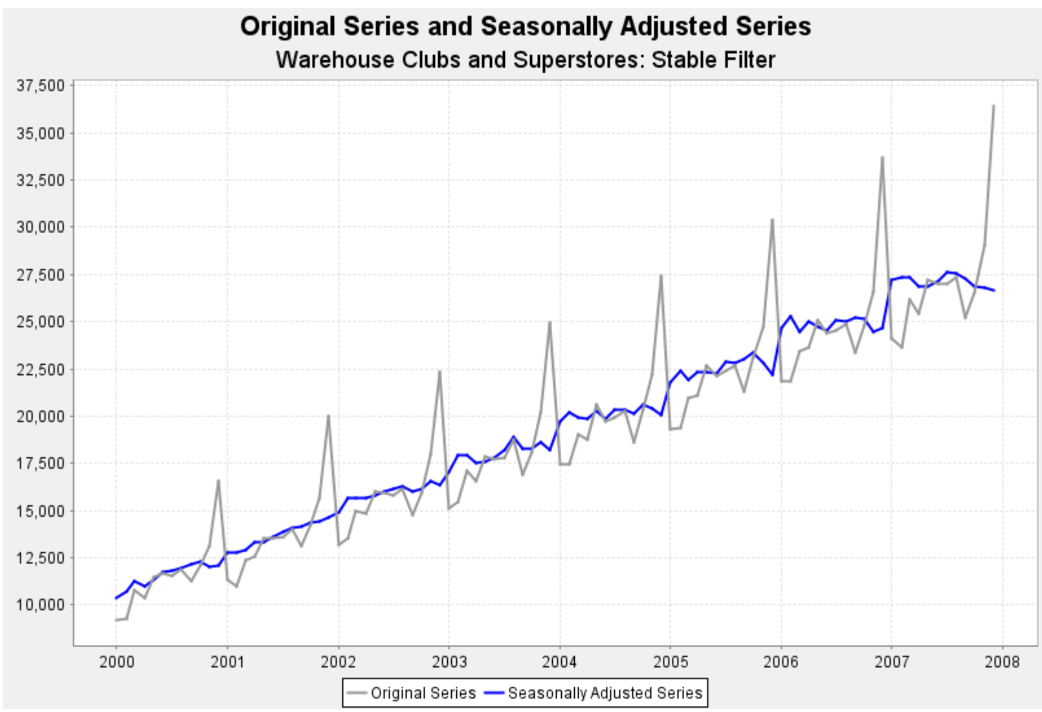


Figure 1: The graph shows the last eight years of the original series overlaid with the last eight years of the the stable-factor seasonal adjustment of the 16 year data span that starts in January 1992. Both *arspec* and *pdg* detect residual seasonality in the last eight years of the adjusted series, see Figures 2 and 3.

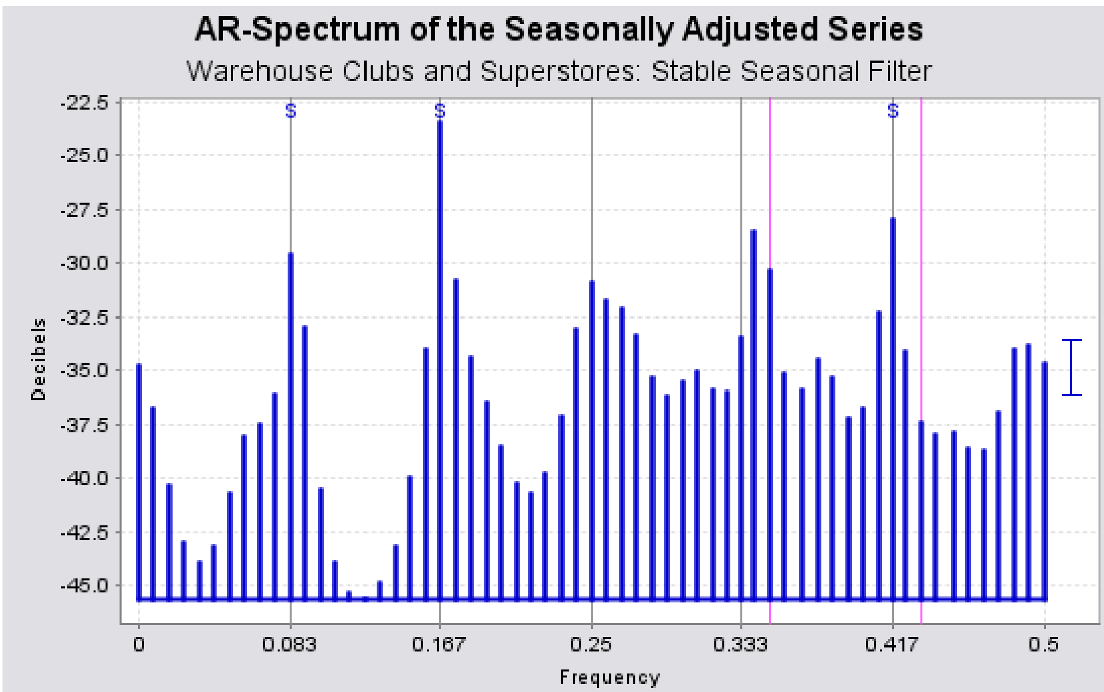


Figure 2: The scale bar at the far right shows the decibel amplitude required for a peak to be v.s. The S above the peaks at the first, second and fifth seasonal frequencies indicates visual significance. Because there are multiple v.s. seasonal peaks, this is classified as a *strong arspec* detection of residual seasonality in the seasonal adjustment span shown in Figure 1, see Table 3. (The vertical lines at .348 and .432 identify the locations of trading day frequencies.)

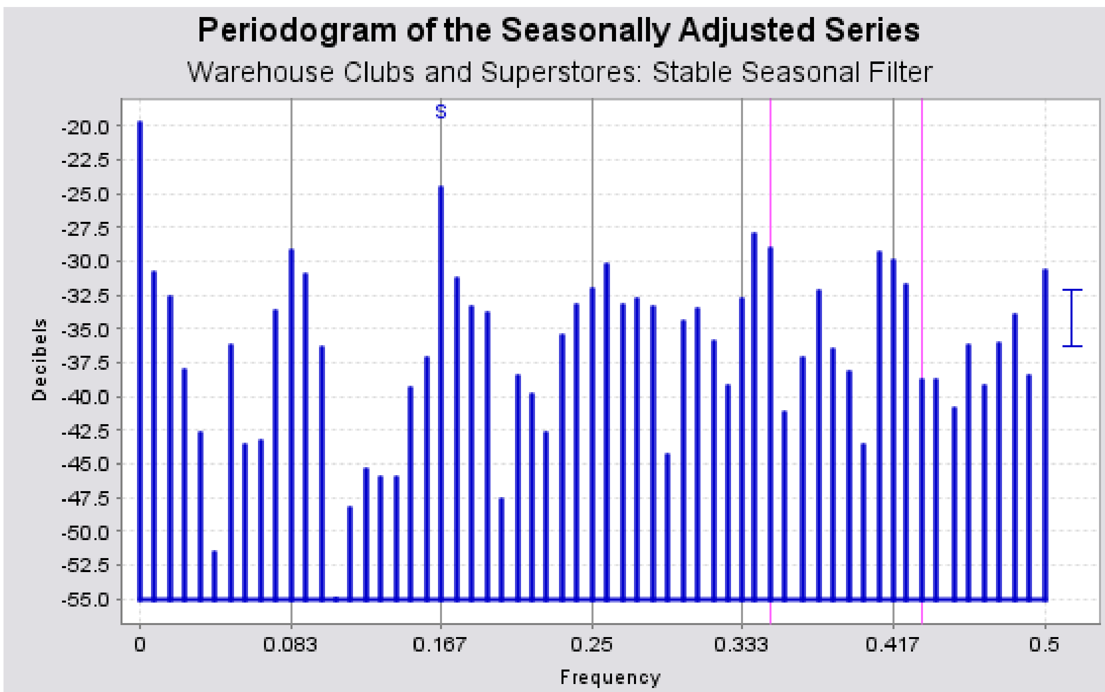


Figure 3: *pdg* detects residual seasonality in the final 8 years of the seasonal adjustment in Figure 1 but with one only v.s. peak, one of the three of *arspec*'s in Figure 2.

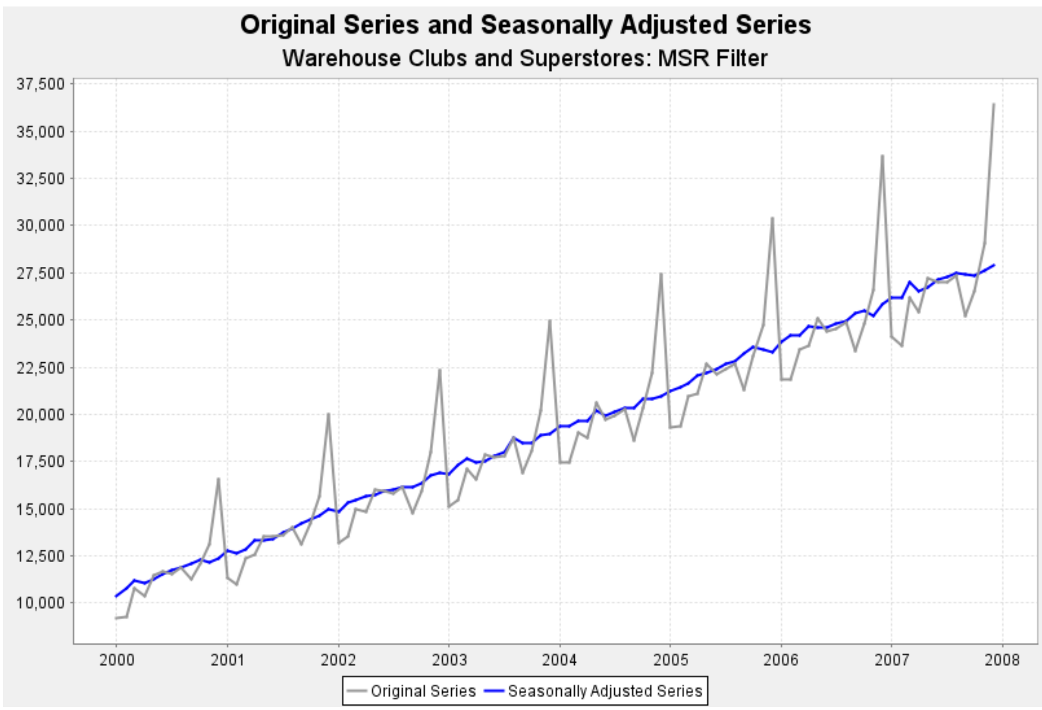


Figure 4: Automatic filter choice of a 3x5 seasonal filter results in a smooth seasonally adjusted series with no *arspec* or *pdg* indications of residual seasonality, as Figure 5 shows for *arspec.*

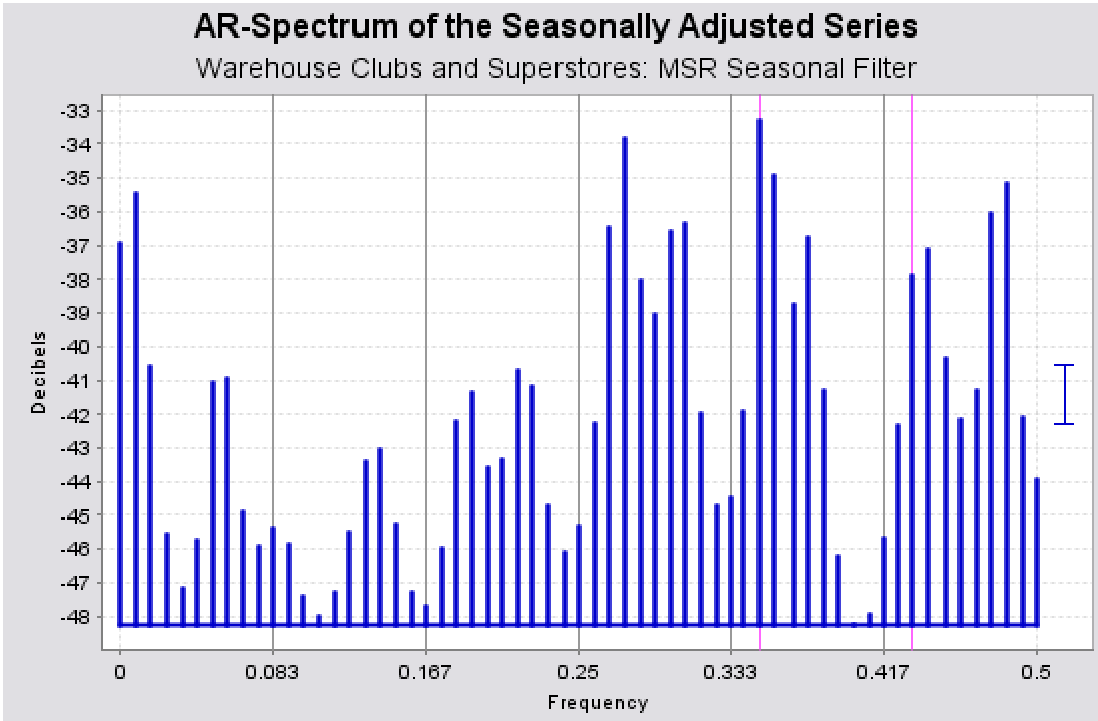


Figure 5: Automatic X-11 filter choice results in the seasonally adjusted series whose last eight years are shown in Figure 4. For these years neither *arpec* nor *pdg* (not shown) has a v.s. peak.

## 3. The RegARIMA GLS F-Statistic for Stable Seasonality

For determining if a regARIMA modeled seasonal time series has statistically significant stable seasonality, the stable-seasonal $F^M$ statistic was developed in Lytras *et al.* (2007) to take advantage of the Generalized Least Squares (GLS) regression coefficient estimates provided by X-12-ARIMA and its successors. $F^M$ can be calculated for any regARIMA model whose regression component includes the stable-seasonal regressors defined in (14) below and whose ARIMA differencing polynomial does not have the seasonal sum factor $1 + B + B^2 + \cdots + B^{11}$. (Otherwise the differencing operation that precedes parameter estimation would zero these regressors.) For example, the ARIMA model cannot have $1 - B^{12} = (1 - B)(1 + B + B^2 + \cdots + B^{11})$.

For $j = 1, \ldots, 11$, the stable-seasonal regressors of X-13A-S, T-S and JD+ are

$$M_{j,t} = \begin{cases} 1 & \text{in month } j \\ -1 & \text{in December} \\ 0 & \text{otherwise} \end{cases} . \tag{14}$$

The stable-seasonal regression function $\Sigma_{j=1}^{11} \beta_j M_{j,t}$ has the alternate constrained form

$$\Sigma_{j=1}^{11} \beta_j M_{j,t} = \Sigma_{j=1}^{12} \alpha_j m_{j,t} \tag{15}$$

in the monthly indicator variables ($m_{j,t} = 1$ in month $j$ and $m_{j,t} = 0$ otherwise) with coefficients

$$\alpha_j = \begin{cases} \beta_j, & 1 \leq j \leq 11 \\ -\Sigma_{j=1}^{11} \beta_j & j = 12 \end{cases} \tag{16}$$

obeying the constraint $\Sigma_{j=1}^{12} \alpha_j = 0$. Because for each $j$, the regressor $m_{j,t}$ sums to one over any 12-month interval, (14) shows that the period 12 regression functions $\Sigma_{j=1}^{11} \beta_j M_{j,t}$ sum to zero over such an interval and over any interval whose length is a multiple of 12. Most often, the coefficients are estimated from log transformed data, and $\exp(\Sigma_{j=1}^{12} \hat{\alpha}_j m_{j,t}) = \Pi_{j=1}^{12} \exp(\hat{\alpha}_j m_{j,t})$ shows that the $j$-th month's estimated stable effect is $\exp(\hat{\alpha}_j)$, with $\hat{\alpha}_j$ the estimate of $\alpha_j$.

Let $\hat{\chi}^2 = \hat{\beta}' \left[ var \left( \hat{\beta} \right)^{-1} \right] \hat{\beta}$ denote the software's chi-square statistic for any specified regression model with $k$ coefficients $\beta = (\beta_1, \ldots, \beta_k)'$ whose estimate $\hat{\beta}$ has been jointly obtained with the ARMA coefficients, e.g. by iterative GLS estimation given an ARIMA (p,d,q)(P,0,Q)$_{12}$ model for the regression disturbances as in Otto, Bell and Burman (1987). See also Galbraith and Zinde-Walsh (1992). For our exposition, the first 11 coefficients of $\beta$ are the coefficients of the $M_{j,t}$, with $j > 11$ reserved for any other appropriate regressors for the empirical series.

Lytras *et al.* (2007) provide simulation evidence that

$$F^M = \frac{\hat{\chi}^2}{11} \times \frac{n - d - k}{n - d},$$

approximately follows an $F^M_{11,n-d-k}$ distribution when there is no stable-seasonal component and the model's other regression and ARIMA specifications can describe the mean function and autocovariances of the data. (For seasonal periods $s \neq 12$, e.g. $s = 4$ with quarterly data, $s - 1$ replaces 11 in the $F^M$ formula.) Lytras *et al.* (2007) test at the .05 level. Other F-tests, e.g. for $M$-$T$, use the .01 level. We use $F_{\alpha}^{M}$ to indicate the test at level $0 < \alpha < 1$ and similarly for F-tests of other diagnostics.

As mentioned in the Introduction, Bell (2011) shows that residual seasonality in a seasonally adjusted series can only occur when the seasonality in the original series is changing over time and that, regardless of the strength of the moving seasonality, for the full span of the seasonally adjusted series, the $F^M$ statistic's null hypothesis of no stable seasonality will be correct (approximately with an X-11 multiplicative adjustment).

However, we will show empirically that $F^M$ can be used to detect residual seasonality if it is calculated from a sufficiently reduced subspan of the adjusted series. Often the last eight years of a somewhat longer series suffices. The following subsection provides an informal explanation.

### 3.1. What do Stable-Seasonal Regression Coefficients Estimate?

When the seasonal regression function $\Sigma_{j=1}^{12}\alpha_j m_{j,t}$ is estimated from log-transformed data as in our empirical study, the text below (16) described how the $\exp(\hat{\alpha}_j)$ define stable seasonal factors. With the series from Subsection 2.4, we first show graphically that each $\exp(\hat{\alpha}_j)$ can approximate the "mean" (sample average) value of the $j$-th calendar month's estimated time varying seasonal factors, both those of the X-11 multiplicative adjustment and those of the SEATS log-additive adjustment of this series.

For the 16 year span January 1992-December 2007, Figure 6 shows each calendar month's changing multiplicative X-11 seasonal factors and their means. The factors result from the automatically chosen $3 \times 5$ seasonal filter. Figure 7 shows that each mean is close to (i) the month's X-11 stable-seasonal adjustment factor, (ii) the month's $\exp(\hat{\alpha}_j)$ from the stable-seasonal GLS regression estimate $\hat{\alpha}_j$ with an automatically chosen ARIMA model, and (iii) the average of the month's SEATS model-based seasonal factor estimates from this model. See Subsection 7.2 for a formal treatment of calendar month means and their fundamental properties.

Figure 7 indicates that the stable-factor adjustment of the 16 year span will shift the calendar month factors shown in Figure 6 in an order-preserving way to have means close to 1.0. The log transformation used prior to regression estimation transforms factors and calendar month means above 1.0 to positive numbers, those below 1.0 to negative numbers. The interpretation of the $e^{\hat{\alpha}_j}$ derived from Figure 7 also suggests that the statistically significant $\hat{\alpha}_j$ will have the signs of these transformed quantities.

One can test this interpretation by applying it to the last halves of the calendar month graphs in Figure 6, i.e. to the last eight years of X-11 seasonal factors produced with use of the $3 \times 5$ filter. The interpretation of the coefficient estimates given above suggests that calendar months whose seasonal factor estimates change most in these final 8 years should have the most significant estimates $\hat{\alpha}_j$ from this span, positive if the change is an increase, negative if a decrease. From Table 2, the reader can decide how well the t-statistics of the $\hat{\alpha}_j$ conform to these expectations.

**Table 2. t-Statistics of the $\hat{\alpha}_j$ of the 8-year span ($|t_j| > 2$ in bold)**

| $j$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t_j$ | 1.77 | **4.78** | **3.24** | 0.97 | 0.6 | 0.46 | 1.56 | 1.24 | -1.11 | **-2.04** | **-4.39** | **-7.01** |

The full seasonally adjusted series has no stable seasonality, so with longer subspans, the mean of each calendar month's seasonal factors along with each $\exp(\hat{\alpha}_j)$ will move closer to 1.0, and the number of $\hat{\alpha}_j$ that are significantly different from zero will become negligible, see the 12-year span results in Table 5 below.

A reader interested in a formal mathematical analysis with such conclusions can consult Subsections 7.2 and 7.3 of the Appendix.

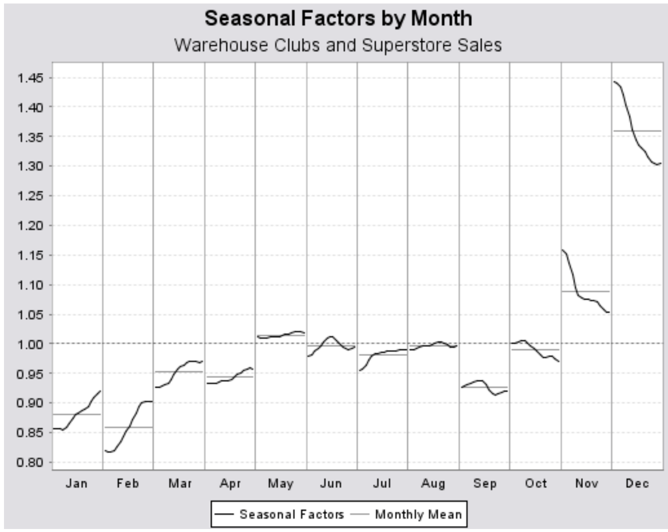


Figure 6: Multiplicative seasonal factors by calendar month for Sales of Warehouse Clubs and Superstores, January 1992–December 2007. The horizontal lines show the means. The factors are those from default X-11 seasonal filter selection.

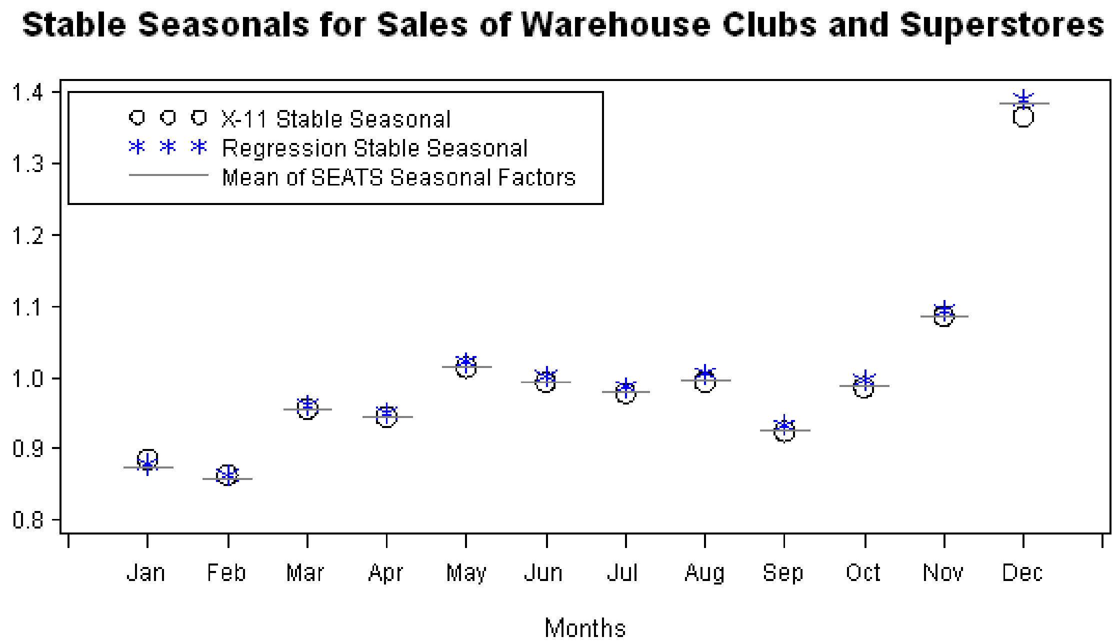


Figure 7: This graph reveals that the exponentiated GLS stable seasonal coefficient estimates $\exp(\hat{\alpha}_j)$ are, effectively, estimates of their month's average seasonality. They are very close to the calendar month sample means of both the log-additive SEATS seasonal factors and the multiplicative X-11 stable seasonal factors of Figure 6.

## 4. Maravall's QS Statistic for Positive Seasonal Autocorrelation

The $QS$ statistic of Maravall (2012) for detecting *positive* seasonal autocorrelation is a function of the first and possibly also the second seasonal-lag sample autocorrelation of the stationary transform of the data being investigated. The stationarized values $x_t$ are obtained by applying $(1-B)^d$ to the data or their logs with appropriate $0 \leq d \leq 2$. For $x_t$ from the original series or from seasonally adjusted data, $d = 2$ if the ARIMA model of the unadjusted series has both a seasonal and a nonseasonal differencing. Otherwise, initially $d = 1$, but the program changes this to $d = 2$ if a check indicates that this is preferable. For irregulars $d = 0$.

With $c_k, k \geq 0$ as in (9) and $r_k = c_k/c_0$, and with $s$ denoting the seasonal period, $s = 12$ here, the focus is on the signs of the sample seasonal autocorrelations, $r_l, l = s, 2s$.

Following Maravall (2012), let $nz$ denote the length of the undifferenced data span and set $n = nz - d$. Define

$$R_l = \begin{cases} r_l, & \text{if } r_l > 0 \\ 0, & \text{if } r_l \leq 0 \end{cases}.$$

When $R_s = 0$, set $QS = 0$. Otherwise set

$$QS = n(n+2)\left\{\frac{R_s^2}{n-s} + \frac{R_{2s}^2}{n-2s}\right\}.$$

As an approximation motivated by simulations described in Maravall (2012), the statistic $QS$ is taken as having a $\chi^2(2)$ distribution under the hypothesis $\rho_l \leq 0$, $l = 1, 2$. Testing is done at the .01 level of $\chi^2(2)$, so the test is denoted by $QS_{.01}$.

*Remark.* There is always positive probability, perhaps small, that $QS = 0$, i.e. that a sample autocorrelation estimate will be negative. Hence $QS$ does not have a continuous density function such as a chi-square. Thus the specified .01 level cannot be taken literally. Self and Liang (1987) illustrate the kinds of distributions that can arise in this situation. The important fact is that $QS_{.01}$ performs well in our empirical study.

### 4.1. Background for QS with Residual Seasonality

A detection of $\rho_{12} > 0$, positive seasonal lag autocorrelation, either in the stationarized monthly seasonally adjusted series or in the irregulars, is a strong indication of residual seasonality. It is strong because $r_{12} < 0$, suggesting $\rho_{12} < 0$, is what is expected after seasonal adjustment. The negative $r_{12}$ phenomenon for stationarized seasonally adjusted values is the focus of the article McElroy (2012) and its references. The seasonal adjustments of all but one of the 88 series of the article's empirical study have $r_{12} < 0$, with statistical significance for 46 at the .05 level (two-sided null hypothesis, see p. 41 of McElroy (2012) for an example indicating the test's calculations). Similarly 86 of the 88 irregular components have $r_{12} < 0$, with statistical significance for 59. (For the tests, the estimated model coefficients were assumed to be correct.)

An intuitive explanation, given a formal foundation in Findley, Lytras and Maravall (2016), hereafter Findley et al. (2016), is that removal of estimated seasonal factors, which generally evolve smoothly from year to year within each calendar month as in Figure 6 above, results in *nonsmooth* consecutive movements in each calendar month's stationarized seasonal adjustment and also in its irregulars, leading to $r_{12} < 0$ in both cases. (Consecutive same calendar month values are 12 months apart on the time scale of the observed series.) The referenced formal foundation links $\rho_{12} > 0$, positive autocorrelation between consecutive values, with smoothness and $\rho_{12} < 0$ with nonsmoothness.

Here we summarize results obtained from the theoretical case of series whose models are completely known and whose unobserved components are estimated from bi-infinite data. We reexpress the basic results, calculated for the semiannual case $s = 2$ of the seasonal random walk in Table 1 of Findley et al. (2016), in terms of the monthly case $s = 12$. First, for the stationarized unobserved seasonal adjustment component $(1-B)^d SA_t$ and its estimated seasonally adjusted series $(1-B)^d \widehat{SA}_t$, we have

$$0 \geq \rho_{12}^{(1-B)^d SA} > \rho_{12}^{(1-B)^d \widehat{SA}}. \tag{17}$$

In the terminology of Findley et al. (2016), (17) establishes that each calendar month series of the stationarized seasonal adjustment $(1-B)^d \widehat{SA}_t$ is *more nonsmooth* than the corresponding unobserved calendar month series of $(1-B)^d SA_t$.

Similarly, for the irregular component $u_t$ and its estimate $\hat{u}_t$, we have

$$0 = \rho_{12}^{u} > \rho_{12}^{\widehat{u}}. \tag{18}$$

Thus $\rho_{12}^{\widehat{u}}$ is negative, so each calendar month series of $\hat{u}_t$ is *more nonsmooth* than the corresponding calendar month series of the unobserved white noise $u_t$.

We have also verified the properties (17) and (18) for airline model components and their estimates for the representative airline model coefficient pairs shown in Tables 7 and 8 of Findley et al. (2016).

Such results make clear why empirical findings of $r_{12}^{(1-B)^d \widehat{SA}} < 0$ and $r_{12}^{\widehat{u}} < 0$ are expected after an appropriate seasonal adjustment, which in turn makes clear why $\rho_{12} > 0$ detections strongly indicate residual residual seasonality.

## 5. Empirical Residual Seasonality Detections

We turn to results for sixteen U.S. Census Bureau Monthly Retail Trade Survey series. All are series for which a seasonal model indicative of moving seasonality was selected for the 16 year data span January 1992–December 2007 by X-13A-S's implementation of a recent version of the automatic ARIMA model selection procedure of T-S (a refinement of the procedure of Gómez and Maravall (2001)). Also for half of the series, the automatic X-11 filter selection option selected the very short $3 \times 3$ seasonal filter for this span, a filter that tends to produce rapidly changing seasonal factors, instead of the alternative longer $3 \times 5$ filter, whose seasonal factors tend to change less. The seasonal adjustment is performed on the series adjusted for regression modeled trading day, holiday, and outlier effects, i.e., the X-11 B 1 series, which we call the *unadjusted series.*

The series titles listed below are preceded by their NAICS codes (with trailing zeroes added) and followed by an indication if the $3 \times 3$ filter was selected.

| | |
|---|---|
| 44000 | Retail and food services sales, total |
| 44300 | Electronics and appliance stores ($3 \times 3$) |
| 44312 | Computer and software stores |
| 44400 | Building materials and garden equipment and supplies dealers |
| 44510 | Grocery stores |
| 44800 | Clothing and clothing accessory stores ($3 \times 3$) |
| 44811 | Men's clothing stores |
| 44812 | Women's clothing stores ($3 \times 3$) |
| 44820 | Shoe stores ($3 \times 3$) |
| 45100 | Sporting goods, hobby, book, and music stores. |
| 45200 | General merchandise stores ($3 \times 3$) |
| 45210 | Department stores - excluding leased departments ($3 \times 3$) |
| 45291 | Warehouse clubs and superstores |
| 45400 | Nonstore retailers ($3 \times 3$) |
| 45410 | Electronic shopping and mail-order houses ($3 \times 3$) |
| 72200 | Food services and drinking places |

Among series whose codes have the same initial digits, the one with the smallest code is the most aggregate and the rest are subaggregates (also called components) thereof. For example, all series with codes in the range 44300 through 44820 are subaggregates of 44000. Also 44312 is a subaggregate of 44300.

As was mentioned earlier, the X-11 stable-factor adjustment of each series left residual seasonality that was detected in the final eight year span by at least one diagnostic, *arspec* always, and usually by others too. By contrast, when the seasonal adjustment for each full 16 year January 2002-December 2007 series is obtained with the automatic X-11 seasonal filter selection option, there are no significant seasonal peaks or significant $F^{M}_{.01}$ or $F^{fs}_{.01}$values for the the last 8 year span, and no significant $QS_{.01}$ values for the adjustment of the full January 16 year 1992 to December 2007 span. ($QS_{.01}$ detections tend to increase with the span length, as will be seen.)

For the final 8 year subspans, detection results for each diagnostic and each adjusted series and its associated irregular component are provided below in Tables 4 and 6 respectively, noting which are strong detections as defined in Table 3. In the 8 year span results for spectrum diagnostics, especially *arspec*, it will be seen in Tables 4 and 6 that when an aggregate with several subaggregates has two or more significant peaks, then usually one of the subaggregates has a significant peak in common with the aggregate.

For longer subspans, only each diagnostic's detection totals from the sixteen series are provided, in Tables 5 and 7, with separate totals for strong detections.

### 5.1. Detections from the Stable-Seasonal Adjustments

Table 3 defines *strong* detections for each diagnostic. The rest are *weak* detections. Tables 4-5 show seasonally adjusted series results. Table 6-7 show the corresponding irregulars results. The diagnostics are calculated after X-11 extreme value adjustment; see Dagum (1980) or Ladiray and Quenneville (2001). In the Totals row of these tables, the entries in parentheses are for strong detections. With spectral detections, each significant seasonal peak's number is shown. For *M-T,* Maravall (2012) describes simulation-based approximate F-statistic criteria of T-S, also implemented in the other software, for what are called weak and strong peaks for all six seasonal frequencies. Strong *M-T* peaks are indicated by * in Tables 4 and 6. Maravall's detection criteria are always as demanding as those in Table 3 and in several cases more demanding.

**Table 3. Criteria for a Strong Detection**

| Diagnostic | Strong Detection |
|---|---|
| *arspec* | 2 v.s. peaks |
| *pdg* | 2 v.s. peaks |
| *M-T* | One $F_{.01}$ peak (*) or two $F_{.05}$ peaks |
| $F^{fs}$ | .01 significance |
| Stable-seas. GLS $F^M$ | .01 significance |
| $QS$ | .01 significance |

For the case in which each diagnostic is calculated from the last 8 years of the seasonally adjusted series, Table 4 identifies weak and strong detections. The weak detection total for each diagnostic is shown in the final row before its number of strong detections, which is shown in parentheses. Regarding strong detections, *pdg* performs substantially worse than other diagnostics, with only 2 compared to 8 or more for the others.

**Table 4. Detections (Strong Detections) of Residual Seasonality in the Last 8 Years of Stable-Seasonal Adjustments**
**( - indicates no detection)**
*Italic* **codes identify series at most one strong detection.**

| Series | *arspec* | *pdg* | *M-T* | $F_{.05}^{fs}$ $(F_{.01}^{fs})$ | $F_{.05}^{M}$ $(F_{.01}^{M})$ | $(QS_{.01})$ |
|---|---|---|---|---|---|---|
| 44000 | (1,2,4) | 2 | 2 | (.01) | (.01) | - |
| 44300 | (2,4) | - | 2 | (.01) | (.01) | (.01) |
| 44312 | 3 | - | (3*) | .05 | .05 | (.01) |
| 44400 | 2 | 1 | 2 | .05 | (.01) | (.01) |
| *44510* | (1,4) | 1 | - | - | - | - |
| 44800 | (1,2,3,4) | - | - | (.01) | (.01) | - |
| 44811 | (1,2) | 2 | (1*,2*) | .05 | - | - |
| 44812 | (1,3) | - | - | (.01) | (.01) | - |
| 44820 | 4 | 2 | (1,2) | (.01) | .05 | (.01) |
| 45100 | 4 | 4 | - | (.01) | (.01) | (.01) |
| 45200 | (1,2) | - | (2*) | (.01) | (.01) | - |
| 45210 | 2 | - | (1,4) | .05 | (.01) | (.01) |
| 45291 | (1,2,5) | 2 | (1,2*) | (.01) | (.01) | (.01) |
| 45400 | 2 | (1,2) | (2,3) | (.01) | (.01) | (.01) |
| 45410 | (1,2) | (1,2) | (1,2*) | (.01) | (.01) | (.01) |
| *72200* | 1 | 1 | 1 | - | .05 | - |
| Totals | 16 (9) | 10 (2) | 12 (8) | 14 (10) | 14 (11) | (9) |

Seasonality in the adjusted series is detected by at least two diagnostics for all sixteen series. Fourteen have two or more strong detections. For the final 8 years, the regression diagnostics have the most strong detections (as happens again for the irregulars, with even more detections, as Table 6 shows). The weakest results are those for 44510 (Grocery Stores), which only has spectrum detections, and 72200 (Food Services and Drinking Places). These series also have the weakest detection results from irregulars series, see Table 6. From their definitions, one would expect at most modest evolution over time in their seasonal patterns.

For *arspec*, with the exception of 44312, at least one of any subaggregate's v.s. peaks is also a v.s. peak of the largest aggregate. Also, with the exception of 44820, a seasonal peak that is significant for *arspec* is significant for one or more other spectrum diagnostics, for either the seasonally adjusted series or the irregulars, often both. With *M-T*, 44000, 45200 and 45400 have a peak in common with a subaggregate. Also, for ten series, at least one of the frequencies with a significant *M-T* peak also has a v.s. *arspec* peak. It is an attractive feature of spectrum diagnostics that they can reinforce indications of residual season in multiple ways.

Table 5 summarizes residual seasonality detection totals for the final 8, 10 and 12 year spans and also the full 16 year data span. Regarding strong detections, only $QS$ consistently detects the same or more with each increase of interval length, having more than other diagnostics already with the 10 year spans. Counting weak detections, *arspec* has the most detections at all lengths. The diagnostics *pdg* and $F^M$ perform substantially worse with each increase in span length. For $F^M$ this is the result predicted in Subsection 3.1 whose argument also applies to $F^{fs}$.

**Table 5. All Spans Detection (Strong Detection) Summaries for the Stable-Seasonal Adjusted Series**

| span | *arspec* | *pdg* | *M-T* | $F^{fs}_{.05}$ $(F^{fs}_{.01})$ | $F^{M}_{.05}$ $(F^{M}_{.01})$ | $QS_{.01}$ |
|---|---|---|---|---|---|---|
| 8yr | 16 (9) | 10 (2) | 12 (8) | 14 (10) | 14 (11) | (9) |
| 10yr | 15 (9) | 9 (0) | 12 (7) | 8 (3) | 10 (7) | (10) |
| 12yr | 13 (4) | 7 (0) | 2 (0) | 1 (0) | 0 (0) | (10) |
| 16yr | 14 (9) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | (12) |

Both Tables 4 and 5 show that *pdg* (with the v.s. criterion) is inferior to *arspec* and that *pdg* makes no contribution to the number of series in each span with a strong detection. The same holds for the strong irregular series detections, see Tables 6 and 7 below.

$QS_{.01}$ is revealed as an important diagnostic for detecting residual seasonality. Positive seasonal correlation after appropriate differencing usually indicates a deficient seasonal adjustment, see Section 4.

### 5.2. Detections from the Associated Irregulars

Tables 6 and 7 are the analogues for the irregulars of Tables 4 and 5. Numbers in **bold** emphasize the few cases (resp. totals) in which the irregulars detections are more in number or strong more often than the detections from the seasonal adjustment. This happens often enough to show that both the seasonal adjustment and the irregulars should be examined for residual seasonality.

**Table 6. Detections (Strong Detections) of Residual Seasonality in the Last 8 Years of the Irregulars**
**( - indicates no detections)**
*Italic* **codes identify series without multiple strong detections**

| Series | *arspec* | *pdg* | *M-T* | $F^{fs}_{.05}$ $(F^{fs}_{.01})$ | $F^{M}_{.05}$ $(F^{M}_{.01})$ | $(QS_{.01})$ |
|---|---|---|---|---|---|---|
| 44000 | (2,4) | **(1,2)** | **(2*)** | (.01) | (.01) | - |
| 44300 | (2,4) | **-** | 6 | (.01) | (.01) | (.01) |
| 44312 | (**1**,**3**) | - | 3 | .05 | - | (.01) |
| 44400 | (**1**,**2**) | 1 | - | .05 | (.01) | - |
| *44510* | - | - | - | - | - | - |
| 44800 | (1,2,3,4) | - | 2 | (.01) | (.01) | **(.01)** |
| 44811 | 2 | 2 | $(1,2^{*})$ | (.01) | (.01) | **(.01)** |
| 44812 | (1,3) | 1 | - | (.01) | (.01) | **(.01)** |
| 44820 | - | 2 | 2 | (.01) | (.01) | - |
| 45100 | (**1**,**4**) | 4 | 2 | (.01) | (.01) | (.01) |
| 45200 | 2 | - | (2*) | (.01) | (.01) | - |
| 45210 | 2 | - | (1,**2**,4) | (.01) | .05 | (.01) |
| 45291 | (2,5) | 2 | (2*) | (.01) | (.01) | (.01) |
| 45400 | 2 | 2 | (2,3*) | (.01) | (.01) | (.01) |
| 45410 | (1,2) | (1,2) | (2*,3) | (.01) | (.01) | (.01) |
| *72200* | 1 | 1 | - | - | (**.01**) | - |
| Totals | 14 (9) | 10 (2) | 12 (7) | 14 (12) | 14 (13) | (**10**) |

Regarding seasonality detections in the last 8 years of the extreme value adjusted irregulars (X-11 output table E 3), Table 6 shows that, as in Table 4, series 44510 and 72200 have fewer than two strong detections. In Table 6 as in Table 4, for *arspec*, with few exceptions, at least one of any subaggregate's v.s. peaks is also a v.s. peak of the largest aggregate. In Table 6, for nine series, at least one of the frequencies with a significant *M-T* peak also has a v.s. *arspec* peak.

The results for increasing span lengths in Table 7 parallel those of Table 5, whose discussion applies also to Table 7. The least successful diagnostic is again *pdg*.

**Table 7. Summary of Detections in E 3 Irregulars (Strong Detections)**

| span | *arspec* | *pdg* | *M-T* | $F^{fs}_{.05}$ ($F^{fs}_{.01}$) | $F^{M}_{.05}$ ($F^{M}_{.01}$) | $QS_{.01}$ |
|---|---|---|---|---|---|---|
| 8yr | 14 (9) | 10 (2) | 12 (7) | 14 (12) | 14 (13) | (**12**) |
| 10yr | 14 (6) | 5 (0) | 4 (1) | 12 (**10**) | 6 (4) | (**12**) |
| 12yr | 13 (3) | 5 (0) | 2 (**1**) | 1 (**1**) | 0 (0) | **(14)** |
| 16yr | 12 (5) | 1 (1) | 0 (0) | 0 (0) | 0 (0) | (14) |

## 6. Concluding Remarks

Because of its consistently inferior performance, we recommend against use of *pdg*. The diagnostics found to be effective in our empirical study, when properly applied, are implemented, sometimes with minor modifications, in easily available seasonal adjustment programs. Our empirical study and background results should help users of such software to make reliable diagnoses regarding residual seasonality. This document could also be of general interest regarding the nature of residual seasonality.

Institutions that publish seasonal adjustments have differing practices regarding whether the full span for which the adjustment is calculated is published or only a final subspan. In any case, because more recent data are usually of greater interest, seasonal adjustment quality control should include application to a perhaps unpublished subspan (or its irregulars) of a seasonality diagnostic expected to be sensitive to seasonality in such a span according to results like those of our analyses.

Our finding that most diagnostics must be applied to a subspan of the seasonally adjusted series for best residual seasonality detection raises doubts about the heralded model-based seasonal adjustment quality control procedure of checking for *idempotency*, see for example Nardelli (2008). For this procedure, the software is rerun in an automatic mode on the full seasonally adjusted series. If a model decomposition without a seasonal component is produced, and therefore no seasonal adjustment of the seasonally adjusted series is done, there is *idempotency*, meaning no change to the initial seasonal adjustment. This would be an important property if it meant that there is no residual seasonality. But it does not establish that residual seasonality cannot be found in a subspan. Further, in the rare situations when idempotency fails, there is no research showing that automatic seasonal adjustment of the seasonally adjusted series is likely to produce a series with no residual seasonality.

We have only considered monthly series. Lytras (2015) provides results on the performance of quarterly versions of the diagnostics for detecting residual seasonality in quarterly series, simulated and real, including the $F^M_{.05}$ and $QS_{.05}$ and *arspec* diagnostics and X-13A-S diagnostics inherited from its predecessors. Most diagnostics perform poorly because of the small numbers of observations. $QS_{.05}$ is the most credible diagnostic, usually with the full series.

Finally, it is very useful that seasonal adjustment software can be run is such a way that it immediately provides the output of seasonality diagnostics for the adjusted series. But this will not account for post-adjustment modifications made prior to publication. For example, central banks and statistical offices sometimes apply formal or informal benchmarking methods that modify seasonally adjusted series in order to force them to satisfy accounting constraints, e.g., forcing directly seasonally adjusted components of an aggregate series to sum to the direct seasonal adjustment of the aggregate, see Quenneville and Fortier (2012) and den Butter and Fase (1991). Such modified seasonal adjustments need to also be tested for residual seasonality.

*Acknowledgments.* The authors are grateful for detailed comments and helpful suggestions on earlier drafts of this document by William Bell, Katherine McDonald-Johnson, and Brian Monsell.

## 7. Appendix

### 7.1. Inconsistency and Other Properties of $\tilde{g}(\lambda)$

The sample spectral density $\tilde{g}(\lambda)$ of (8) does not converge to $g(\lambda)$ as $n \to \infty$ because it employs sample autocovariances (9) that average too few values of $(x_t - \bar{x})(x_{t+k} - \bar{x})$ to estimate enough $\gamma_k$ adequately, e.g. $c_{n-1} = n^{-1}(x_1 - \bar{x})(x_n - \bar{x})$ estimates $\gamma_{n-1}$ from one term. A revealing convergence result is

$$Var\left(\frac{\tilde{g}(\lambda)}{g(\lambda)}\right) = \frac{Var(\tilde{g}(\lambda))}{g^2(\lambda)} \to \begin{cases} 1, & 0 < \lambda < 1/2 \\ 2, & \lambda = 0, 1/2 \end{cases}, \qquad (19)$$

see §10.3 of Brockwell and Davis (1991). This shows how $\tilde{g}(\lambda)/g(\lambda)$ fails to converge to 1. ($Var(1) = 0$.) It also shows that $Var(\tilde{g}(\lambda))$ is frequency-dependent, changing when $g(\lambda)$ does. Most importantly, it shows that the ratio has the same asymptotic variance for all interior frequencies $0 < \lambda < 1/2$ but double this variance at the endpoint frequencies $\lambda = 0, 1/2$. As (11) suggests, when $\mu = 0$ the same results hold for $pdg(\lambda)$.

The key result motivating log transformed diagnostics, described precisely after formula (13), is somewhat analogous. It shows that taking logs stabilizes the asymptotic error variance, making it constant over all interior frequencies, doubling at endpoint frequencies. This justifies using the same v.s. criterion at all seasonal frequencies other than $\lambda = 1/2$ for *pdg* and the *M-T* diagnostic, and also for *arspec* (apply the proof of Corollary 5.6.3 of Brillinger (1975) to the results of Theorem 6 of Berk (1974)).

*Remark.* Simulation and empirical experiments could help to decide if a 12* or other criterion for v.s. *arspec* peaks at $\lambda = 1/2$ would be effective. Maravall (2012) outlines simulations for the alternative v.s. criterion of T-S but the incomplete discussion does not suggest a simple criterion.

### 7.2. The Calendar-Month-Average Stable Seasonal Component

Suppose $m \geq 1$ years of detrended or stationarized monthly data $x_1, \ldots, x_{12m}$ are considered with $x_1$ from January. For $1 \leq j \leq 12$ and $1 \leq k \leq m$, the $j$-th calendar month's datum in the $k$-th year of data is $x_{j+12(k-1)}$. The 12 calendar month subseries of ,

$$x_{j+12(k-1)}, 1 \leq k \leq m, 1 \leq j \leq 12 \tag{20}$$

have sample means,

$$\overline{x}_j(m) = m^{-1} \Sigma_{k=1}^{m} x_{j+12(k-1)}, 1 \leq j \leq 12, \tag{21}$$

which provide a decomposition of the sample mean of the data,

$$\overline{x}(12m) = (12m)^{-1} \Sigma_{l=1}^{12m} x_l = (1/12) \Sigma_{j=1}^{12} \overline{x}_j(m).$$

The centered calendar month means

$$s_j(m) = \overline{x}_j(m) - \overline{x}(12m), 1 \leq j \leq 12 \tag{22}$$

satisfy $\Sigma_{j=1}^{12} s_j(m) = 0$. Hence a stable seasonal component for the data is given by

$$s_{j+12(k-1)} = s_j(m), 1 \leq k \leq m, 1 \leq j \leq 12. \tag{23}$$

It will be convenient to focus on zero-mean SAR(1) data (4). Then the 12 calendar-month subseries of length $m$ years, $x_{j+12(k-1)}, 1 \leq k \leq m, 1 \leq j \leq 12$, have AR(1) models with coefficient $\Phi$, innovation variance $\sigma_a^2$, and variance $\sigma_a^2 \left(1 - \Phi^2\right)^{-1}$. From (5), they are mutually uncorrelated, so the same is true of their calendar-month sample means $\overline{x}_j\left(m\right)$, whose variance is the same for all $j$, $E\overline{x}_j\left(m\right)^2 = E\overline{x}_1\left(m\right)^2$. We will use these properties to obtain formal results yielding observed properties of $F_{.01}^M$.

### 7.3. Insights from SAR(1) Series

A noted property of $F_{.01}^M$ in Tables 5 and 7 is that its stable seasonality indications are greatest with the shortest subspan, substantially diminishing as the span length increases. The informal results of Subsection 3.1 regarding the stable seasonal component defined by average monthly seasonality suggested explanations. For SAR(1) series (4), we show formally in Subsection 7.3.1 that the standard deviations $\sqrt{Es_j\left(m\right)^2}$ of (22) and (23), which depend on $m$ and on the seasonal-lag autocorrelation $\Phi$, are relatively large for small $m$ (provided $\Phi$ is not too small), but tend to zero at the rate $m^{-1/2}$ as $m \rightarrow \infty$.

Table 8's detection rates for $F_{.01}^M$ from 5000 simulation for each $\Phi$ and $n = 12m$ display the anticipated $F_{.01}^M$ detection changes: decreased detections as $m$ increases or $\Phi$ decreases.

**Table 8. Rates of Strong $F^M_{.01}$ Detections for $\Phi = 0.2, 0.4$ with Increasing $n$**

| $\Phi$ | $n = 12m$ | $F^M_{.01}$ |
|---|---|---|
| .4 | 96 | .376 |
| .4 | 144 | .204 |
| .4 | 288 | .076 |
| .2 | 96 | .363 |
| .2 | 144 | .197 |
| .2 | 288 | .070 |

### 7.3.1. Analysis of the SAR(1) Stable-Seasonal Component

From Section 7.2, $\bar{x}(12m) = (1/12)\Sigma_{j=1}^{12}\overline{x}_j(m)$ is an uncorrelated-component decomposition of $\overline{x}(12m)$ in which each component has the same variance $E\overline{x}_1(m)^2/144$. Therefore (using $\Longleftrightarrow$ for "equivalent formula") its variance and standard deviation satisfy

$$144E\overline{x}(12m)^2 = 12E\overline{x}_1(m)^2 \Longleftrightarrow \sqrt{E\overline{x}_1(m)^2} = \sqrt{12}\sqrt{E\overline{x}(12m)^2}. \qquad (24)$$

For (22), we have $12s_j(m) = 12\overline{x}_j(m) - 12\overline{x}(12m) = 11\overline{x}_j(m) - \overline{x}(12m)/12 = 11\overline{x}_j(m) - \Sigma_{l=1,l\neq j}^{12}\overline{x}_l(m)$, a decomposition into mutually uncorrelated calendar month components. For $1 \leq j \leq 12$, uncorrelatedness and (24) yield

$$Es_j(m)^2 = Es_1(m)^2 = \frac{11}{12}E\overline{x}_1(m)^2 \Longleftrightarrow \sqrt{Es_j(m)^2} = \sqrt{11}\sqrt{E\overline{x}(12m)^2}. \qquad (25)$$

Thus, when $\sqrt{E\overline{x}(12m)^2}$ is non-negligible, the magnitudes $\sqrt{s_j(m)^2}$ of the stable-seasonal component (23) can be expected to be more so. The analysis of Subsection 3.1 suggests that $F^M_{.01}$ will indicate statistically significant stable seasonality, which Table 8 confirms. However, for increasing $m$, Theorem 8.3.1 of Anderson (1971) shows that $\lim_{m\to\infty}\sqrt{12m}\sqrt{E\{\overline{x}(12m)\}^2} = \sqrt{g(0)} = \sigma_a(1-\Phi)^{-1}$. Thus for the stable-seasonal components, one has $\sqrt{E\{s_j(m)\}^2} \to 0$ at a rate proportional to $(12m)^{-1/2}$.

Concerning dependence on $\Phi$, for fixed $m$, it can be calculated for any zero-mean stationary series $x_t$ with autocovariances $\gamma_j$ that

$$144mE\overline{x}(12m)^2 = \gamma_0 + 2\left\{\Sigma_{j=1}^{m-1}\left(1-\frac{j}{m}\right)\gamma_j\right\}.$$

For fixed $m$, since $\left|\gamma_j\right| \leq \gamma_{0,}$ the right hand expression tends to zero if $\gamma_0 \rightarrow 0$, which means $\Phi \rightarrow 0$ in the SAR(1) case, so the same is true of $\sqrt{E\overline{x}\left(12m\right)^2}$ and $\sqrt{E\left\{s_1\left(m\right)\right\}^2}$.

## References


[1] Akaike, H. and M. Ishiguro. 1980. BAYSEA, *A Bayesian Seasonal Adjustment Program*, Computer Science Monographs No. 13 Tokyo: The Institute for Statistical Mathematics.

[2] Anderson, T. W. 1971. *The Statistical Analysis of Time Series.* New York: Wiley.

[3] Bell, W. R. 2011. "Unit Root Properties of Seasonal Adjustment and Related Filters." Statistical Research Division Research Report *Statistics #2010-08.* (Revised August 30, 2011) Washington DC: U.S. Census Bureau. Available at: http://www.census.gov/ts/papers/rrs2010-08.pdf (accessed January 24, 2016).

[4] Berk, K. N. 1974. "Consistent autoregressive spectral estimates." *Annals of Statistics* 2: 489-502.

[5] Brillinger, D. R. 1975. *Time Series: Data Analysis and Theory.* New York: Holt Rinehart and Winston.

[6] Brockwell, P. J. and Davis, R. A. 1991. *Time Series: Theory and Methods, 2nd Edition.* New York: Springer Verlag.

[7] Caporello, G. and A. Maravall. 2004. *Program TSW: Revised Manual*, Documentos Ocasionales 0408, Bank of Spain.

[8] Dagum, E. B. 1980. *The X-11-ARIMA Seasonal Adjustment Method*, Statistics Canada. http://www.census.gov/ts/papers/1980X11ARIMAManual.pdf (accessed June 3, 2016)

[9] De Antonio, D. and J. Palate. 2014. JD+ Seasonality Tests, Brussels: National Bank of Belgium, (December 16, 2014 draft).

[10] Den Butter F. A. G. and M. M. G. Fase. 1991. Seasonal Adjustment as a Practical Problem, Amsterdam:North-Holland.

[11] Findley, D. F., D. P. Lytras, and A. Maravall. 2016. "Illuminating ARIMA Model-Based Seasonal Adjustment with Three Fundamental Seasonal Models," *SERIEs* 7,11-52. DOI: 10.1007/s13209-016-0139-4 http://link.springer.com/article/10.1007/s13209-016-0139-4 (accessed February 26, 2016).

[12] Galbraith, J. W. and V. Zinde-Walsh. 1992. "The GLS Transformation Matrix and a Semi-Recursive Estimator for the Linear Regression Model with ARMA Errors" *Econometric Theory* 8: 95–111.

[13] Gómez, V. and A. Maravall. 1996. *Programs TRAMO and SEATS: Instructions for the User* (beta version: June 1997). Madrid: Banco de España, Servicio de Estudios, DT 9628.

[14] Gómez, V. and A. Maravall. 2001. "Automatic Modeling Methods for Univariate Series." In *A Course in Time Series Analysis.* edited by D. Peña, G. C. Tiao, and R. S. Tsay, 1–42. New York: Wiley and Sons.

[15] Ladiray, D. and B. Quenneville. 2001. *Seasonal Adjustment with the X-11 Method.* Lecture Notes in Statistics, Vol. 158 New York: Springer.

[16] Lytras, D. P. 2015. "On Seasonality: Comparing X-13ARIMA-SEATS Diagnostics on Quarterly Series." In 2015 Proceedings of the Business and Economic Statistics Section: American Statistical Association. Alexandria, VA: American Statistical Association. Available at http://www.census.gov/ts/papers/lytras2015.pdf (accessed December 4, 2015).

[17] Lytras, D.P., R. M. Feldpausch, W. R. Bell. 2007. "Determining Seasonality: A Comparison of Diagnostics from X-12-ARIMA," *Proceedings of the Third International Conference on Establishment Surveys* (Montreal, June 2007). Available at http://www.census.gov/ts/papers/ices2007dpl.pdf (accessed May 1,2013).

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

[19] Marple, S. L. Jr. 1987. *Digital Spectral Analysis with Applications.* New York: Prentice-Hall.

[20] McElroy, T. S. 2012. "An Alternative Model-Based Seasonal Adjustment That Reduces Residual Seasonal Autocorrelation." *Taiwan Economic Forecast and Policy* 43: 33-68. http://www.econ.sinica.edu.tw/english/content/periodicals/list/2013093010104847832/

[21] McElroy, T. S. 2008. "Matrix Formulas for Nonstationary ARIMA Signal Extraction." *Econometric Theory* 28, 9881009. Preprint available at http://www.census.gov/ts/papers/matform3.pdf

[22] Nardelli, S. 2008. "Seasonal Adjustment Quality Reports" Ch. 8 of *Seasonal Adjustment* (eds. M. Manna and R. Peronaci) Frankfurt am Main: European Central Bank https://www.ecb.europa.eu/pub/pdf/other/statseasonaladjustmenten.pdf (accessed November 6, 2014).

[23] National Bank of Belgium. 2015. https://www.nbb.be/en/publications-and-research/econometric-software (accessed November 2015).

[24] Newton, H. J. and M. Pagano. 1983. "A Method for Determining Periods in Time Series." *Journal of the American Statistical Association* 78: 152-157.

[25] Otto, M. C., W. R. Bell, and J. P. Burman. 1987. An iterative GLS approach to maximum likelihood estimation of regression models with ARIMA errors. *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, 632-637. Available at https://www.census.gov/ts/papers/rr87-34.pdf (accessed December 20).

[26] Priestley, M. 1981. *Spectral Analysis and Time Series*. London: Academic Press.

[27] Quenneville, B. and Fortier, S. 2012. "Restoring Accounting Constraints in Time Series–Methods and Software for a Statistical Agency." In *Economic Time Series: Modelling and Seasonality*, edited by W. R. Bell, S. H. Holan and T. S. McElroy) Boca Raton: CRC Press, Taylor and Francis Group.

[28] Self, S. G. and K. Y. Liang. 1987. "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions," *Journal of the American Statistical Association* 82: 605-610.

[29] 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. *1999 Proceedings of the American Statistical Association*, Business and Economic Statistics Section, pp. 144-149. Alexandria, VA: American Statistical Association. http://www.census.gov/ts/papers/rr9903s.pdf (accessed March 1, 2016).

[30] U.S. Census Bureau. 2016. *X-13-ARIMA-SEATS Reference Manual*, *Ver. 1.1.* http://www.census.gov/ts/x13as/docX13ASHTML.pdf (accessed March 1,2016).

[31] Wikipedia Contributors. 2012. Complex Number, *Wikipedia, The Free Encyclopedia.* http://en.wikipedia.org/wiki/Complex_number (accessed January 29, 2012).