

A Review of Seasonal Adjustment Diagnostics

Tucker McElroy¹  and Anindya Roy^{1,2} 

¹U.S. Census Bureau, Suitland, MD, USA

²University of Maryland, Baltimore County, Baltimore, MD, USA

Email: tucker.s.mcelroy@census.gov

Summary

Seasonal adjustment methods are used to process and publish thousands of time series across the world each month, and judgement of the adequacy relies heavily upon seasonal adjustment diagnostics. This paper discusses tests for the adequacy of a seasonal adjustment, first reviewing the broader background on tests for seasonality and then proceeding to four tests that are appropriate for stationary forms of seasonality. We contrast time-domain and frequency-domain approaches, focusing upon four diagnostics available in the seasonal adjustment literature (and software packages) and applying the methods to a large collection of public use time series. Each of the four tests is designed around a distinct formulation of seasonality, and hence, empirical performances differ; we compare and contrast the methods and include discussion on how diagnostic results can be used to improve faulty seasonal adjustments. Directions for future research, involving inverse partial autocorrelations and polyspectra, are also discussed, and the methodologies are supported by theoretical and simulation results.

Key words: QS diagnostic; seasonal autocorrelation; spectral peaks; visual significance.

1 Introduction

Quarterly or monthly economic time series typically exhibit seasonality, most often described via a non-stationary stochastic process with unit root frequencies corresponding to the known seasonal frequencies (Bell & Hillmer, 1983); also see the discussion in chapter 3 of Hylleberg (1986). Because the presence of seasonality can interfere with the interpretation of a time series (as seasonality can mask trend movements, for example), it has become standard practice at governmental statistical agencies to adjust time series, that is, estimate and extract the seasonality (see the discussion in Bell & Hillmer, 1984, 1985). If after the application of a seasonal adjustment procedure there is residual seasonality present, then the adjustment is inadequate. Hence, testing for residual seasonality is a problem of widespread importance; thousands of time series are seasonally adjusted each month at statistical agencies around the world, many of whom utilise the software program X-13ARIMA-SEATS (X-13ARIMA-SEATS Reference Manual, 2015) of the U.S. Census Bureau. The detection of seasonality in raw (or unadjusted) data is a different problem from that of testing for residual seasonality in adjusted data, as the target specification and statistical properties of the time series are different in the two cases. This paper is focused on the latter case (testing for residual seasonality); we

begin with a discussion of raw seasonality, review the available tests in the literature and finally focus upon the more narrow problem of residual seasonality.

1.1 Fundamentals of Seasonality

We begin by discussing the empirical phenomenon of seasonality and develop some mathematical concepts that parse the phenomenon; we do not entertain a single mathematical framework, lest such a definition proves to be too narrow, thereby excluding important cases. Fairly recent discussions of the phenomenon of seasonality can be found in Fase *et al.* (1973), Bell & Hillmer (1984), den Butter & Fase (1991), Franses (1996) and Findley (2005), although interest in the subject goes back at least to Buys Ballot (1847). Conceptually, seasonality is persistency¹. The term ‘persistency’ is used to denote a high degree of association in a stochastic process between various random variables, possibly measured through autocorrelation, spectra or fixed effects (regressors). The exact definition will depend on whether the stochastic process is stationary or non-stationary, and the form the persistency takes. In a time series over seasonal periods that is not explainable by intervening time periods. Behaviour that recurs annually (i.e. over the seasonal period) is an example; however, if the persistence of behaviour from year to year is driven by a persistence between the seasons, then the dynamic is not truly seasonal—it may instead be arising from trending effects. If the seasonality is sufficiently strong, it can be assessed as a fixed effect or possibly also through seasonal unit roots (discussed further in the succeeding text; also see Proietti, 1996). Weaker persistence in the seasonality can be assessed through autocorrelations and spectral peaks, both of which are facets of a stationary stochastic process.

We begin with two examples designed to illustrate a continuum of behaviour from weak seasonality to strong seasonality—see Findley *et al.* (2015) for additional discussion. In the following, let $\{X_t\}$ be a weakly stationary time series with autocovariance function $\gamma_h = \text{Cov}[X_{t+h}, X_t]$, and let $\rho_h = \text{Corr}[X_{t+h}, X_t]$ be the autocorrelation function.

Example A seasonal process. With s corresponding to the number of seasons (so $s = 12$ for monthly data), consider the seasonal autoregression (SAR) defined via

$$X_t = \Phi_1 X_{t-s} + \epsilon_t,$$

where $\{\epsilon_t\}$ is i.i.d. with mean zero and variance σ^2 . To ensure stationarity of $\{X_t\}$, set the parameter Φ_1 to take values in the set $(-1, 1)$, with larger positive values. The value $\Phi_1 = 1$ corresponds to a non-stationary process, the SAR has a zero frequency unit root and $s-1$ seasonal unit roots. Assuming $h \geq 0$, the autocorrelation function (acf) is equal to $\Phi_1^{|h|}$ if $h = ns$ for some integer n and is zero otherwise. Hence, if Φ_1 is large, there is high correlation at seasonal lags, and there is no correlation at non-seasonal lags. Figure G.1 of Appendix G in the supporting information depicts various values of the autocorrelation function, in the quarterly case $s = 4$. Simulations, corresponding to each value of Φ_1 , are depicted in Figure G.2 of Appendix G in the supporting information, to show the visual correspondence. As the parameter Φ_1 increases from 0.6 to 0.8, 0.9 and 0.99, the structured seasonal pattern is more apparent, indicating a lag 4 association. The autocorrelation plots reflect this behaviour, with weak seasonality indicated by $\Phi_1 = 0.6$ and strong seasonality apparent in the case of $\Phi_1 = 0.99$.

To see why association at seasonal lags is not sufficient to describe seasonality, suppose that correlation was high at non-seasonal lags. Then it could happen that the high seasonal

correlation is generated through the presence of high season-to-season correlation. The next example develops this idea explicitly.

Example A non-seasonal process. Consider the regular autoregression (AR), defined via

$$X_t = \phi_1 X_{t-1} + \epsilon_t,$$

where $\{\epsilon_t\}$ is i.i.d. with mean zero and variance σ^2 . The parameter ϕ_1 takes values in $(-1, 1)$ to ensure stationarity of $\{X_t\}$, with larger positive values indicating a stronger persistency. The autocorrelation function is $\rho_h = \phi_1^{|h|}$. Hence, $\rho_s = \phi_1^s$, which could be large if ϕ_1 is close to unity. Although there would then be a high seasonal lag correlation, the correlation at non-seasonal lags is also high, and this nullifies the seasonal effect. Figure G.3 of Appendix G in the supporting information depicts various values of the autocorrelation function, in the quarterly case $s = 4$. Simulations, corresponding to each value of ϕ_1 , are depicted in Figure G.4 of Appendix G in the supporting information, to show the visual correspondence. The values of ϕ_1 are chosen such that the autocorrelation at seasonal lags exactly matches the values from the SAR; that is, ϕ_1 increases from 0.880 to 0.946, 0.974 and 0.997. Clearly, the simulations indicate no seasonal lag associations, and the behaviour of the autocorrelation plot is quite different from that of the SAR.

In order to account for season-to-season correlation, we might consider the partial autocorrelation at a seasonal lag, defined via $\kappa_h = \text{Corr}[X_{t+h}, X_t | X_{t+1}, \dots, X_{t+h-1}]$. Here, the vertical bar denotes linear projection—see McElroy & Politis (2020) for background. The following example, however, shows that the partial autocorrelation also fails as a device to measure seasonality, because some truly seasonal processes will be deemed non-seasonal.

Example A cyclic process. Consider the order 2 autoregression defined via

$$X_t = 2\rho \cos(2\pi/s) X_{t-1} - \rho^2 X_{t-2} + \epsilon_t,$$

where $\{\epsilon_t\}$ is i.i.d. with mean zero and variance σ^2 . The parameter $\rho \in (-1, 1)$ measures the persistency of the cyclic behaviour, which has frequency $2\pi/s$, or period s . The autocorrelation function is $\rho_h = a_1(\rho e^{i2\pi/s})^{|h|} + a_2(\rho e^{-i2\pi/s})^{|h|}$, where

$$a_1 = \frac{\rho^2 e^{-i2\pi/s} - e^{i2\pi/s}}{(1 + \rho^2)(e^{-i2\pi/s} - e^{i2\pi/s})} \quad a_2 = \frac{e^{-i2\pi/s} - \rho^2 e^{i2\pi/s}}{(1 + \rho^2)(e^{-i2\pi/s} - e^{i2\pi/s})}.$$

As a result, $\rho(s) = \rho^s$, which is necessary for seasonality; Figure G.5 of Appendix G in the supporting information depicts various values of the autocorrelation function in the quarterly case $s = 4$, with corresponding simulations displayed in Figure G.6 of Appendix G in the supporting information. Evidently, the process is seasonal when ρ is sufficiently high, but $\kappa_h = 0$ for $|h| > 2$, and in particular, $\kappa_s = 0$. Hence, partial autocorrelation at the seasonal lag fails to characterise seasonality.

We note that the inverse autocorrelations also suffer the same deficiencies as the partial autocorrelations for characterising the cyclic process, but the inverse partial autocorrelations show some promise (see Appendix A in the supporting information); however, there is no available published research on such a seasonality diagnostic tool. In summary, the concept of persistency (from year to year) alone is not sufficient to describe seasonality, but it is necessary. If the process is stationary, persistency can be directly measured through the autocorrelation function; or, if non-stationary in the mean, through sinusoidal regressors—so that seasonality corresponds to a fixed effect. If the seasonality is dynamic (i.e. is evolving over time), then fixed effects are

insufficient to capture the phenomenon, and instead, we may have recourse to non-stationary processes, such as seasonal unit root processes.

The frequency domain—examined through the spectral density (defined in the succeeding text)—is another useful way to measure seasonality, as it allows us to assess annual persistency while screening out cases like Example 2. The essential idea is that it may be possible to characterise seasonality in terms of autocorrelations that are high at seasonal lags and much lower at other nearby lags—visually, this quality could be described as an oscillatory behaviour in ρ_h of period s . These sorts of oscillations, as shown in McElroy (2021), correspond to peaks in the spectral density; hence, it follows that seasonality can be characterised by peaks in the spectral density. Similarly, seasonality can be defined through the distribution of autoregressive roots using a partial fraction decomposition, as in Brewer *et al.* (1975), Brewer (1979), Piccolo (1982) and Proietti (1995).

1.2 Measuring and Detecting Seasonality

Year-to-year persistency in a process can be articulated through autocorrelations, as discussed in the previous subsection, or through a periodic deterministic function, such as a sinusoid. One approach to detection is to postulate as an alternative hypothesis the existence of a sinusoid—corresponding to a deterministic seasonal component—at the frequency of interest, and test whether spectral density estimates warrant such a hypothesis. The early literature on spectral peak (SP) testing (see Priestley, 1981, pp. 390–415) focused on this approach, whereby one attempts to detect a fixed seasonal effect embedded in stationary noise. The null hypothesis corresponds to a stationary process without a jump in the spectral distribution, whereas the alternative is a stationary process with some deterministic seasonality. Alternatively, one can test for the presence of such sinusoids in the time domain, by using generalised least squares. This is equivalent to the approach of the model-based F (MBF) test of Lytras *et al.* (2007), where the seasonality corresponds to part of the process' mean (and is represented by seasonal dummy regressors), the remainder of the process corresponding to a specified ARIMA model (Bell, 2004).

Actually, this type of approach can be traced back at least to Buys Ballot (1847), which re-expresses the time series in a vector representation of s annual series and tracks the extent to which the s means differ; also see Chen & Fomby (1999). Thinking of this as a regression problem, it follows that when the errors are i.i.d., we can test for discrepancies in these seasonal means via an ANOVA approach. If we utilise a non-parametric framework, we obtain the Kruskal–Wallis (KW) test for the presence of fixed seasonality (Kruskal & Wallis, 1952). Similarly, the stable seasonality (SS) test is a variant based on using rankings of observations rather than actual values (Friedman, 1937) and is typically applied to so-called SI ratios, which are just a form of de-trended time series data (see Dagum, 1980; Ladiray & Quenneville, 2012). A serious drawback of the KW and SS tests is that the regression errors are presumed to be i.i.d., which very few time series are likely to satisfy. This deficiency is rectified in the MBF test, where the regression errors can now be serially correlated and can even be non-stationary.

Non-stationary seasonality is a natural generalisation of SS (i.e. using seasonal regressors), because such non-stationary processes can include such a stable component without loss of generality—this is analogous to the fact that a random walk with drift can be decomposed into a linear term (its mean) plus the purely stochastic mean zero portion. Tests for stable sinusoids focus on the deterministic part of seasonality but are not designed to address the stochastic portion (except as a sort of nuisance parameter, so to speak). However, in many cases, it is important to account for the stochastic portion of seasonality, as this is accounting for the time-varying facets of the seasonal dynamics (see Hylleberg, 1986). For most economic series,

the seasonality is too evolutive to be adequately captured by fixed periodic functions, and the approach of testing for seasonal unit roots (Gregoir, 2006) has been quite popular. For instance, Proietti (2000) considers a seasonal unit root approach to defining and modelling seasonality.

The HEGY test (Hylleberg *et al.*, 1990) posits a seasonal unit root null hypothesis and a stationary seasonality alternative. In contrast, the Canova–Hansen (CH) test (see Canova & Hansen, 1995) tests a null hypothesis of deterministic seasonality, where the alternative hypothesis corresponds to a time-varying regression that is equivalent to a seasonal unit root formulation. The BH test of Buseti & Harvey (2003) can be adapted to the case of zero seasonality (as a null hypothesis), with either deterministic or unit root seasonality as the alternative. The MBF test is in a sense reversed from the HEGY test, in that the null hypothesis is stationary seasonality, whereas the alternative is deterministic seasonality; in the MBF test, the null cannot include a seasonal unit root (only non-seasonal forms of non-stationarity are allowed).

Another procedure is that proposed by Franses (1994), based on the vector representation (or ‘stacking’) developed by Gladyshev (1961), Tiao & Grupe (1980), Osborn (1988, 1991), Franses & Paap (2004) and Lin *et al.* (2020). This amounts to a test of periodic integration (PI) (a generalisation of seasonal unit roots) as a null hypothesis, with either stationary or seasonal unit root alternatives, depending on the types of co-integration that are tested; a fuller discussion of these possibilities, as well as a treatment of the impact of seasonal adjustment on a periodic time series, can be found in Ghysels & Osborn (2001) and del Barrio Castro & Osborn (2004).

If instead the time series is stationary (as in the case of testing for residual seasonality, after trend differencing), then we may instead have recourse to methods based on either assessing peaks in the spectral density at seasonal frequencies or measuring periodic behaviour in the autocorrelation function. Pierce (1976, 1979) looked at the adequacy of seasonal adjustment by examining the magnitude of the autocorrelations at seasonal lags of the adjusted series. This approach is similar in spirit to that of the Q_s statistic, adopted in TRAMO-SEATS (Maravall, 2012), essentially being a variant of the Box–Ljung–Pierce test applied to seasonal lag autocorrelations. Early work on assessing the effect of seasonal adjustment appeared in Nerlove (1964) and Grether & Nerlove (1970), adopting a frequency-domain approach via examination of the spectral density; this is connected to the spectral approach of Priestley (1981) mentioned earlier.

More specifically, we can obtain insight into seasonality by considering the spectral distribution function and its derivative, the spectral density (Brockwell & Davis, 2013). Whereas a deterministic sinusoid corresponds to a level shift in the spectral distribution function—and will appear in a spectral density estimate as a tall slender peak—stochastic stationary seasonality instead corresponds to a broader peak in a spectral density estimate (e.g. computed using an autoregressive estimator). It will have a broader peak that nonetheless is approaching an infinite height as sample size increases—see discussion in Findley *et al.* (1998). Hence, we can view a process consisting of a sinusoid in the presence of stationary noise as the limiting end of a continuum of processes, whereby stationary seasonality is becoming increasingly pronounced (this is discussed further in the next subsection).

It is important to consider the width of a spectral peak in its assessment. Since the procedure of seasonal adjustment, viewed in the frequency domain, amounts to multiplication of a function with a peak by another function with a trough, the reduction of the peak depends on the width of these functions (McElroy, 2012). For this and other reasons, Soukup & Findley (1999) considered a measure—called visual significance (VS)—of the peak that involved the distance between ordinates of the log spectrum when examined on a grid of frequencies of mesh size $\pi/60$. The VS test considers some mild degree of stationary seasonality as a null hypothesis,

which is rejected in favour of an alternative—corresponding to a higher seasonal persistence—when there is a large peak in the spectral density. (Mild and strong seasonality, in the context of the frequency domain, refers to whether the spectral density at a seasonal frequency has a pronounced peak; a more specific discussion is given in Section 2.4.) Alternatively, McElroy & Holan (2009) proposed a spectral peak diagnostic that assessed the first and second derivatives of the spectral density—referred to as the spectral convexity (SC) test.

A different approach, based on the time domain, is to describe seasonality through an oscillatory autocorrelation function, with higher values occurring at seasonal lags. The ROOT diagnostic of McElroy (2021) examines the oscillatory behaviour of the Wold coefficients of a time series and defines varying degrees of seasonal persistence according to whether the autoregressive representation has a root corresponding to a seasonal frequency. The rationale is that such oscillatory patterns are necessary—and thus processes such as Example 2 can be screened out—as seen in Examples 1 and 3.

This is a fairly exhaustive list of reputable, documented statistical methods, which break into stationary and non-stationary cases. The various tests for seasonality all utilise at least one of these three frameworks: deterministic, stationary or non-stationary. These are not mutually exclusive categories (as noted, a seasonal unit root formulation of non-stationarity always has a deterministic facet as well, without loss of generality). We summarise the various tests in Table 1.

We mention in passing some other methodologies that can be found on the Internet, but lack a statistical foundation, and hence will not be further discussed in this article. The software JDemetra+ has available a frequency domain diagnostic based upon summing the values of the periodogram at seasonal frequencies—when scaled by the total sum, one obtains the coefficient of determination of the trigonometric regression of the series on sines and cosines at the seasonal frequencies. (A referee suggested that a slight modification of the software’s output could result in a suitable diagnostic.) SEATS has a spectrum diagnostic based on applying the Blackman–Tukey Hanning window and assessing spectral peaks by comparing values to the sum of neighbouring values—this is in the spirit of the VS test and adopts as null hypothesis that the time series is white noise. Finally, there are the software-based methods, whereby a series is passed into software such as SEATS or X-12-ARIMA and is deemed seasonal if the automatic model selection has chosen an ARIMA model that includes seasonal differencing. Such a method implicitly relies on the automatic modelling capabilities of the software; see Maravall (2012) and Findley *et al.* (1998) respectively for discussions pertinent to SEATS and X-12-ARIMA.

Table 1. *Tests for seasonality.*

| <i>Test</i> | <i>Null</i> | <i>Alternative</i> |
|-------------|-----------------------------|---|
| BH | Zero seasonality | Deterministic seasonality or seasonal unit root |
| CH | Deterministic seasonality | Seasonal unit root |
| HEGY | Seasonal unit root | Stationary seasonality |
| KW | Stationary seasonality | Deterministic seasonality |
| MBF | Stationary seasonality | Deterministic seasonality |
| PI | Periodically integrated | Stationary seasonality or seasonal unit root |
| Q_s | Zero seasonality | Stationary seasonality |
| ROOT | Strong seasonal persistence | Mild seasonal persistence |
| SC | Zero seasonality | Stationary seasonality |
| SP | Stationary seasonality | Deterministic seasonality |
| SS | Stationary seasonality | Deterministic seasonality |
| VS | Mild seasonal persistence | Strong seasonal persistence |

1.3 The Problem of Residual Seasonality

Seasonal adjustment seeks to remove seasonality in a time series without affecting the non-seasonal dynamics. If seasonality is deterministic, then seasonal adjustment is simply achieved by subtracting the appropriate regressors, multiplied by estimates of the regression parameters. Otherwise, when stochastic facets of seasonality are present, regression techniques are insufficient and one typically proceeds to estimate (or extract) the random seasonal effects. Most commonly, linear filters are employed to do the extraction—this is true whether the seasonality is of the non-stationary or stationary type—and may be constructed from fitted models. Whichever method is used, a successful seasonal adjustment must balance two considerations: all the apparent seasonality has been extracted, and no non-seasonal effects have been extracted.

In order to assess the success of a seasonal adjustment, it is therefore necessary to gauge whether all the seasonality has been removed and whether or not non-seasonal effects have been erroneously removed. Additional criteria for the adequacy of a seasonal adjustment have been proposed; see den Butter & Fase (1991). Because this review is focused more narrowly on the problem of residual seasonality, we address this facet in the succeeding text—our concern is whether all the seasonality has been extracted, and we review the available tests for the presence of residual seasonality in seasonally adjusted series. The problem of whether non-seasonal effects have been extracted (sometimes called ‘over-adjustment’) is discussed in the succeeding text.

Testing for residual seasonality is a different task from testing for seasonality in a raw time series. This is because the properties of such series tend to be quite different. First, it is rarely the case that either deterministic seasonality or non-stationary seasonality is present in a seasonally adjusted series, because most seasonal adjustment procedures are designed so that they always remove both fixed seasonal effects and any seasonal unit root dynamics as well.² By including in the seasonal adjustment filter a polynomial in the lag operator B of the form $1 + B + \dots + B^{s-1}$, where s is the seasonal period, all period s functions will be annihilated by the filter, and any seasonal unit root process will be rendered stationary; see remark 3.5.9 of McElroy & Politis (2020). (An exception occurs when time series are adjusted indirectly, that is, by obtaining the seasonal adjustment by summing up seasonally adjusted dis-aggregated component series—but even in this scenario, any residual seasonality is unlikely to be non-stationary; see discussion in McElroy, 2018.)

Second, any time series already processed by a seasonal adjustment procedure is modified in certain ways that disrupt stationarity; for instance, it is typical to modify seasonal adjustment filters at the beginning and end of a time series (perhaps via forecast extension, as in X-11-ARIMA Dagum, 1980), resulting in different weights and therefore introducing local non-stationarity and non-linearity.

The degree of stationary seasonality that is present in a seasonally adjustment time series is the determinant of adequacy: if this stationary seasonal effect is highly persistent—as in the case of $\Phi_1 = 0.9$ in Example 1—then seasonality will still be apparent in the time series, but if the persistence is low enough, then the adjustment can be deemed adequate. Most practitioners do not require a total absence of such seasonality (this would be like requiring $\Phi_1 = 0$ in a SAR(1)) for adequacy, although some other practitioners consider any amount of stationary seasonality (even as high as $\Phi_1 = 0.99$ in a SAR(1)) to be adequate. Our own view is that residual seasonality diagnostics should be flexible, allowing each practitioner to set their own thresholds of persistence and test accordingly.

Given these considerations, the BH, CH, HEGY, KW, MBF, PI, SP and SS tests are not appropriate for testing for residual seasonality. This leaves the VS, SC, Q_s and ROOT tests, which may be viewed as frequency domain (VS and SC) and time domain (Q_s and ROOT) efforts to

capture the degree of seasonal persistence. We next proceed to discuss these diagnostics in greater detail and their connection to the concept of seasonal persistence. If the acf is absolutely summable, then the spectral density can be defined via $f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_h e^{-ih\lambda}$, for any $\lambda \in [-\pi, \pi]$. Because the autocovariance function is symmetric in h , the spectral density is real and can be expressed in terms of cosines. Supposing that a stationary time series exhibits seasonality, we know that γ_h is high for h of the form ns for n integer, because high correlation at seasonal lags is a way of metrising the notion of seasonal persistence. (However, this is a necessary but not sufficient condition for seasonality.) The cosine function is largest when its argument is equal to a multiple of 2π , so λ of the form $2\pi/s$ or integer multiples of such correspond to $\cos(ns\lambda)$ having a large value, and the spectral density will have a local maximum at such frequencies so long as γ_h has lower values at non-seasonal lags. Specifically, the spectral density will have a peak close to each seasonal frequency $2\pi/s$ when the autocorrelation function is high at seasonal lags and much lower at non-seasonal lags. Such a structure in the acf corresponds to seasonality, because the high autocorrelation at seasonal lags is not obfuscated by high correlation at neighbouring intervening seasons. For instance, in Example 1, the spectral density is equal to

$$\frac{\sigma^2}{|1 - \Phi_1 e^{-i\lambda s}|^2} = \frac{\sigma^2}{1 + \Phi_1^2 - 2\Phi_1 \cos(\lambda s)}.$$

The maximiser is the frequency $2\pi/s$ or any integer multiple thereof. The frequencies $2\pi j/s$ for $j = 1, 2, \dots, s$ are known as seasonal frequencies, because they refer to phenomena of period s/j , which occur a total of j times within the year. In contrast, the spectral density for Example 2 is

$$\frac{\sigma^2}{|1 - \phi_1 e^{-i\lambda}|^2} = \frac{\sigma^2}{1 + \phi_1^2 - 2\phi_1 \cos(\lambda)},$$

which has a maximiser at frequency zero. This is not a seasonal process, and the spectral density has no peaks at seasonal frequencies. For Example 3, the spectral density is equal to

$$\frac{\sigma^2}{|1 - 2\rho \cos(\omega) e^{-i\lambda} + \rho^2 e^{-i2\lambda}|^2} = \frac{\sigma^2}{(1 - 2\rho \cos(\omega - \lambda) + \rho^2)(1 - 2\rho \cos(\omega + \lambda) + \rho^2)}.$$

Consider the case where $\rho \approx 1$; then the maximisers occur for $\lambda \approx \pm\omega$, so that when $\omega = 2\pi k/s$ for integer k , the process is seasonal.

Non-stationary processes can be formulated that generalise the stationary case by having unit correlation at certain lags. In other words, there is a full association between X_t and X_{t-s} . Heuristically, the autoregressive equation

$$X_t = \Phi X_{t-s} + \epsilon_t$$

featured in Examples 1 and 2 can be generalised by allowing $\Phi = 1$; if we maintain $\sigma^2 > 0$, then the process will be stochastic (and non-stationary), but if we shrink σ^2 to zero, then the process will be deterministic. In the stochastic case, application of seasonal differencing yields

$$(1 - B^s)X_t = X_t - X_{t-s} = \epsilon_t,$$

and $\{\epsilon_t\}$ is a stationary process. In the deterministic case, the same equation holds, but now $\epsilon_t = 0$; that is, seasonal differencing annihilates the process. Actually, a deterministic process can still be stationary, though the autocovariances need no longer be absolutely summable, and the spectral density does not exist—instead, we have recourse to the spectral distribution function. The

spectral distribution function can be estimated consistently from stationary time series data, and jump discontinuities correspond to periodic (e.g. seasonal) effects. Whereas a diagnostic based upon the spectral distribution function could be devised, there appears to be no published research on this approach.

For the remainder of the paper, we focus on the stationary case, for which the acf and spectral density are well defined. Based on the ideas of association, if seasonality is present, then ρ_s must surpass a suitably high threshold τ . This definition presumes a stationary time series, because otherwise ρ_s is not well defined. Clearly, setting $\tau = 0$ sets a very low bar—recall from Figures G.1 and G.2 of Appendix G in the supporting information that any value of $\Phi_1 < 0.8$ would correspond to a quite weak seasonal autocorrelation. Also, note that a high value of ρ_s does not necessarily entail the presence of seasonality—recall the discussion of Example 2. On the other hand, a low value of ρ_s does indeed preclude the possibility of seasonality being present. This suggests adopting as null hypothesis

$$H_0: \rho_s = \tau \quad (1)$$

with τ taken fairly large,³ Based on the analysis of the SAR process in Findley *et al.* (2015), a conservative choice is $\tau = 0.7$, because lower values of seasonal autocorrelation imply a negligible degree of autocorrelation after 4 years. and utilising a lower one-sided test such that significant rejections indicate that seasonality is not present. Observe that taking an upper one-sided test will result in fallacious conclusions: if we were to reject the null (say, with a low τ) of little to no seasonality due to a high value of the estimated covariance, we cannot conclude that seasonality is present—because non-seasonal processes can have high ρ_s . In summary, we can use seasonal autocovariances to test whether seasonality is *not* present, but cannot reliably use them to test whether seasonality *is* present; such an endeavour necessitates examination of autocorrelation at other lags or use of the partial autocorrelation function. In order to more fully motivate the frequency-domain approach, we expand on Example 1, examining some seasonal processes to further demonstrate the connection between acf and spectral density.

Example Seasonal peaks. A seasonal process may indeed manifest a peak in the spectral density at a seasonal frequency, as in the process of Example 3. For $\theta = 2\pi/6$, $\rho = 0.85$ and $\sigma = 1$, the acf and spectral density are displayed in the left and right panels of Figure 1. The wave-like form of the acf, with persistence indicated at lags that are a multiple of the seasonal period $s = 12$, is consistent with our definition of seasonality. Also, the frequency $2\pi/6$ in the definition of the AR(2) polynomial corresponds to the second seasonal frequency, and the spectral density exhibits a peak at this

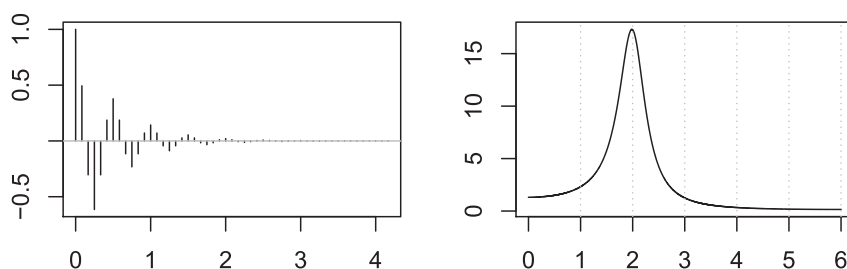


Figure 1. Autocorrelation function (left panel) and spectral density (right panel) for a seasonal AR(2) process. The x-axis of the left panel is lag, expressed in units of years. The x-axis of the right panel is occurrences per year (or frequency divided by $2\pi/12$)

frequency. In contrast to Example 3, consider a more complicated seasonal process, wherein the superposition of seasonal effects slightly shifts the spectral peaks:

$$(1 - 2\rho\cos(\theta_1)B + \rho B^2)(1 - 2\rho\cos(\theta_2)B + \rho B^2)X_t \sim \text{WN}(0, \sigma^2),$$

where $\theta_1 = 2\pi/6$ and $\theta_2 = 3\pi/6$, and $\rho = 0.85$. Here, notice that there are two cyclical portions making up the AR(4) polynomial. The acf and spectral density are displayed in the left and right panels of Figure 2. Here, the oscillatory structure of the acf has been disrupted, and there is only a mild increase in autocorrelation at lag 12 relative to the near neighbours—an indication of weakened seasonality. On the other hand, the spectral density exhibits two peaks somewhat interior to the actual seasonal frequencies $2\pi/6$ and $3\pi/6$. Hence, by considering two mild seasonal effects, the spectral peaks are shifted slightly and the acf has a weaker seasonal pattern. By increasing ρ (not shown), the peaks shift towards the seasonal frequencies and the acf has a more noticeable oscillatory structure. This illustrates the necessity and sufficiency of the spectral peak concept as characterising seasonality, that is, a strong peak indicates an oscillatory acf (and hence seasonality), and conversely, a peak that is either weak or off-shifted from the seasonal frequency corresponds to an acf that lacks strong seasonal oscillations.

Visual significance—which has now become somewhat of a standard by virtue of its incorporation into the X-12-ARIMA software used at many international statistical agencies—is motivated by Examples 3 and 4 and is computed (detailed in Section 2) by comparing the spectral peak ordinate to both nearest neighbour ordinates with respect to the chosen frequency grid; when both ordinate differences exceed a threshold (selected based upon empirical criteria), the spectral peak is declared to be ‘visually significant’. The SC method uses an alternative metric for peaks based on an integrated measure of slope and convexity, and Q_s is based on the behaviour of autocorrelations at seasonal lags. The ROOT method goes further into the shape of the autocorrelation function, requiring an oscillatory pattern to be present in order to screen out spurious indications of seasonality.

1.4 Over-adjustment of Time Series

Over-adjustment of a time series refers to a seasonal adjustment process that has removed not only the seasonality but some of the non-seasonal content as well. Because of the nature of seasonal adjustment filters, the phenomenon of over-adjustment is manifested as spectral troughs at the seasonal frequencies—or equivalently, to negative seasonal autocorrelations (sometimes

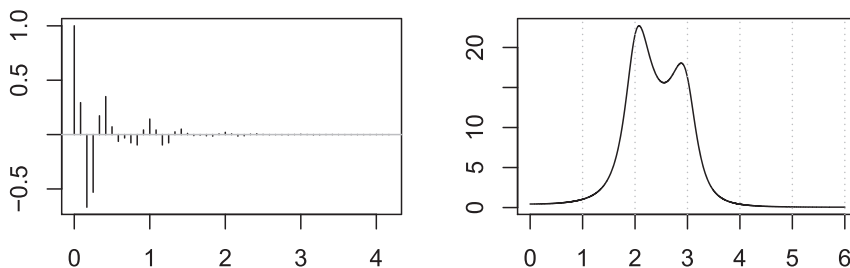


Figure 2. Autocorrelation function (left panel) and spectral density (right panel) for a seasonal AR(4) process. The x-axis of the left panel is lag, expressed in units of years. The x-axis of the right panel is occurrences per year (or frequency divided by $2\pi/12$)

called seasonal anti-persistence). Concerns about over-adjustment motivated the model-based signal extraction diagnostics of A. Maravall (Maravall, 2003; McElroy, 2008; Blakely & McElroy, 2017); however, these diagnostics presume the ARIMA model-based framework and amount to a class of model goodness-of-fit tests focused on a subset of the entire frequency band $[-\pi, \pi]$.

In order to formulate why over-adjustment (and under-adjustment) can be considered pernicious, we must articulate why seasonal adjustment (or more generally, signal extraction) is performed in the first place. Data may be conceived of as being composed of signal (the portion we deem informative or interesting) and noise (the portion that is distracting and interferes with inference), such as when we remove outliers (a form of noise) to obtain extreme-value adjusted data (the signal, in this case). In order to correctly capture the signal, we must not only formulate a framework (or model) of the signal's properties but also must do the same for the noise. Without such frameworks, we would not be able to tell whether we succeeded in extracting all the signal or alternatively whether some of the noise has been extracted into the signal. In the context of seasonal adjustment, where the noise is the seasonal component, allowing some noise into the extracted signal means there is residual seasonality in the seasonal adjustment. Likewise, failing to extract all the signal means that some elements of the data that we are really interested in—such as the trend or business cycle—are wrongfully assigned to the seasonal component and discarded.

The available set of diagnostics for over-adjustment are quite limited—there are the signal extraction diagnostics of A. Maravall, and a version of the ROOT diagnostic adapted to assess the presence of anti-persistence (McElroy, 2021). The reason for this dearth seems to stem from the recognition that optimal filtering (i.e. minimum mean squared error signal extraction) always produces seasonal adjustments with spectral troughs, because this is an intrinsic feature of such filters; see Nerlove (1964), Granger (1978), Sims (1978), Tukey (1978) and Bell & Hillmer (1984). From this point of view, many practitioners deem that over-adjustment is not a problem at all, merely a curious artefact of optimal signal extraction. Of course, such optimal filters are based on fitted models, and the error in parameter estimation (and model misspecification) can result in some degree of under-adjustment or over-adjustment.

Two avenues for research and development arise from the consideration that over-adjustment is not desirable. First, one can design seasonal adjustment filters (or procedures) that do not generate such spectral troughs in the first place—this is the approach of McElroy (2012), which extends the basic idea proposed by Ansley & Wecker (1984). Second, one can design new diagnostics that are focused on flagging over-adjustment. In contrast to diagnostics of under-adjustment, which examine seasonal persistence and seasonal spectral peaks, we would instead examine seasonal anti-persistence and seasonal spectral troughs. Each of the four seasonal adjustment diagnostics considered in this paper can be adapted to focus on over-adjustment: Q_s could be modified so as to focus on negative seasonal autocorrelations, and the ROOT diagnostic could examine roots in the moving average, rather than autoregressive, representation (this has actually been implemented in McElroy, 2021). For the frequency domain, SC could be modified to look at troughs by changing the alternative hypothesis (to concave up), and VS could be restructured to examine trough functionals. Much of the theoretical development discussed in the succeeding text could be easily applied to the adapted diagnostics.

2 Methodology for Assessing Residual Seasonality

We now focus on describing the Q_s , ROOT, SC and VS methods in greater detail, as these are the four published methodologies for assessing residual seasonality.

2.1 The Q_s Diagnostic

If the data have a stochastic or deterministic trend, then it must be differenced to stationarity first; this involves certain perils described in Section 2.5. Let X_1, \dots, X_T denote the resulting sample of a stationary process $\{X_t\}$ that we supposed has been correctly differenced. Then the sample autocovariance at lag s is defined as

$$\hat{\gamma}_s = T^{-1} \sum_{t=1}^{T-s} (X_t - \bar{X})(X_{t+s} - \bar{X}),$$

where $\bar{X} = T^{-1} \sum_{t=1}^T X_t$. Then the sample autocorrelation is $\hat{\rho}_s = \hat{\gamma}_s / \hat{\gamma}_0$. The asymptotic theory of this estimator is given in Brockwell & Davis (2013); it is asymptotically unbiased for ρ_s (as $T \rightarrow \infty$ with s fixed) and is asymptotically normal under fairly broad conditions. The remaining terms in the asymptotic variance depend on other autocovariances and can be determined by plug-in estimators.

The Q_s diagnostic of Maravall (2012) is based upon the preceding discussion of seasonal autocorrelation, adopting $\tau = 0$ in (1) and proceeding with the fallacy that large values of ρ_s and ρ_{2s} must indicate seasonality. In fact, large values of ρ_s and ρ_{2s} in no way preclude a non-seasonal process (cf. Example 2). In order to focus upon upper one-sided alternatives, the sample autocorrelations are squared—but only the positive part; that is, if $\hat{\rho}_s$ or $\hat{\rho}_{2s}$ are negative, then these portions are discarded. If $\hat{\rho}_s \leq 0$, then define Q_s to be zero, but otherwise the formula is

$$Q_s = T(T+2) \left(\frac{\max\{0, \hat{\rho}_s\}^2}{T-s} + \frac{\max\{0, \hat{\rho}_{2s}\}^2}{T-2s} \right), \quad (2)$$

where T is the sample size (after differencing, as needed, to remove trend non-stationarity in the series). The rationale for focusing on positive values of ρ_s and ρ_{2s} appears to be that the diagnostic aims to focus upon residual seasonality only, disregarding the possible presence of so-called anti-seasonality manifested by negative seasonal autocorrelations (which can arise from over-adjustment). The Q_s diagnostic adopts a null hypothesis that the lag s and $2s$ autocorrelations are non-positive, that is,

$$H_0: \rho_s \leq 0 \text{ and } \rho_{2s} \leq 0.$$

The normalisation of the two terms is intended to balance the two contributions and generate a null distribution closely resembling the χ^2 (with two degrees of freedom) under a variety of null processes. So far, there has been no attempt to provide a rigorous asymptotic theory for Q_s (see discussion in Findley *et al.*, 2017), but in Appendix B in the supporting information, we derive the asymptotic distribution in the case that $\rho_s = \rho_{2s} = 0$. Simulations are provided in Appendix C in the supporting information; although the improved asymptotic theory has better size and power (for the case that $\rho_s = \rho_{2s} = 0$), results deteriorate when the asymptotic variance is estimated, and the χ^2 approximation has similar performance.

The diagnostic's inadequacy is built into its definition, as it fails to account for the second portion of seasonality's definition—namely, that seasonal persistence is not explained by inter-seasonal lag effects. It follows from the earlier discussion on seasonal autocorrelation that Q_s is unable to distinguish genuine seasonality from non-seasonal processes that happen to exhibit high seasonal autocorrelation. For example, applying the test to a simulation of an AR(1) process with $\phi_1 = 0.946$ will likely yield rejection of the null, while providing the false conclusion that the data are seasonal.

2.2 The ROOT Diagnostic

The Wold representation (theorem 7.6.4 of McElroy & Politis, 2020) of $\{X_t\}$ can be expressed as $X_t = \mu + \sum_{j \geq 0} \psi_j Z_{t-j}$ for white noise $\{Z_t\}$. Setting $\psi(z) = \sum_{j \geq 0} \psi_j z^j$, and supposing all its zeroes are outside the unit circle of the complex plane, its reciprocal $\pi(z) = 1/\psi(z)$ is well defined in an annulus around the unit circle. This produces the $\text{AR}(\infty)$ representation of $\{X_t\}$; McElroy (2021) shows that oscillatory patterns in the coefficients π_j of $\pi(z) = 1 - \sum_{j \geq 1} \pi_j z^j$ occur if and only if $\pi(z)$ has a zero of the form $z = \rho^{-1} e^{i\omega}$, where $2\pi/\omega$ is the period of the oscillation and $\rho \in (0, 1)$ is its persistence. A ρ -persistent seasonal effect of frequency ω is said to be present whenever $\pi(\rho^{-1} e^{i\omega}) = 0$. Formulated as a null hypothesis of seasonality, we have

$$H_0(\rho_0): \pi(\rho^{-1} e^{i\omega}) = 0 \text{ has solution } r = \rho_0.$$

This hypothesis is tested by fitting an $\text{AR}(p)$ model (where p is taken to be either large or growing with sample size—e.g. determined by AIC) and using the $\text{AR}(p)$ polynomial as an approximation to $\pi(z)$. Let $\phi(z) = 1 - \sum_{j=1}^p \phi_j z^j$ be the $\text{AR}(p)$ polynomial, and suppose the coefficients ϕ_j are estimated by either Yule–Walker equations or OLS, denoted $\hat{\phi}_j$. Then the test statistic is essentially given by $\sqrt{T} \hat{\phi}(\rho_0^{-1} e^{i\omega})$; because this is complex valued, it is convenient to consider the real and imaginary parts—the squared magnitude is equal to the sum of the squared real and imaginary parts. Formulating this as a Wald statistic, we obtain the test statistic

$$T \left(\mathcal{R} \hat{\phi}(\zeta), \mathcal{I} \hat{\phi}(\zeta) \right)' \left(\frac{1}{2} \begin{bmatrix} \mathcal{R} \Xi \\ \mathcal{I} \Xi \end{bmatrix} \hat{\Gamma}_p^{-1} \hat{V} \hat{\Gamma}_p^{-1} \begin{bmatrix} \mathcal{R} \Xi \\ \mathcal{I} \Xi \end{bmatrix}' \right)^{-1} \left(\mathcal{R} \hat{\phi}(\zeta), \mathcal{I} \hat{\phi}(\zeta) \right),$$

where $\zeta = \rho_0^{-1} e^{i\omega}$, $\Xi = [1, \zeta, \zeta^2, \dots, \zeta^p]$, $\hat{\Gamma}_p$ is an estimate of the p -dimensional Toeplitz covariance matrix, and \hat{V} is an estimate of a covariance matrix V . Details on these quantities, and the generalisation to multiple seasonal frequencies, are given in Appendix D in the supporting information, where it is shown that the test statistic is asymptotically χ^2 . An assumption guaranteeing such an asymptotic convergence is that the process is causal and invertible, with innovations that have zero kurtosis. Simulations in Appendix E in the supporting information demonstrate that size and power for the ROOT diagnostic are favourable when the AR model order is known; however, if we account for the unknown model order using AIC, the diagnostic is somewhat mis-sized. It is recommended that at least 20 years of data will be available.

2.3 The Spectral Convexity Diagnostic

In contrast to Q_s , the VS and SC diagnostics associate a peak in the spectrum—at a frequency λ corresponding to the period of a seasonal phenomenon—with persistency; these two methods are founded upon the vast literature of spectral analysis (Marple, 1987). Akaike (1980), Cleveland & Terpenning (1982) and Akaike & Ishiguro (1983) used spectral density plots to detect seasonality; Cleveland & Devlin (1980), McNulty & Huffman (1989) and Soukup & Findley (2001) used the spectral density to identify trading day (Ladiray, 2012) effects. In related work, Newton & Pagano (1983) used spectral peaks to estimate the frequency (or period) of a dynamic effect.

One of the subtleties of our topic is the very definition of peak, which could include assessments of overall height, steepness and convexity. In the most primal conception of a peak, one considers a value that exceeds all neighbouring values; that is, it is a local maximum. The SC

approach adopts this concept by measuring the first and second derivatives of the spectral density over bands of frequencies. Defining the general functional

$$\Xi_A(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) f(\lambda) d\lambda \quad (3)$$

for some kernel A , SC considers the functionals $\Xi_A(\dot{f})$ and $\Xi_A(\ddot{f})$ for kernels that are centred about a frequency of interest; here, \dot{f} and \ddot{f} denote the first and second derivatives of the spectral density. These kernels are twice continuously differentiable, have compact support on an interval of the form $[\theta - \beta/2, \theta + \beta/2] \subset [0, \pi]$, are symmetric about the axis $\lambda = \theta$ and have vanishing first derivative on $\lambda = \theta \pm \beta/2$. It is shown in McElroy & Holan (2009) that the functionals (3) can be re-expressed in terms of \dot{A} and \ddot{A} , and estimators of these functionals can be constructed using

$$\sum_{|h| < T} \hat{\gamma}_h \zeta_h, \quad (4)$$

where ζ_h is the inverse Fourier transform of either \dot{A} or \ddot{A} , that is, the integral of these integrands multiplied by $e^{ih\lambda}$ and divided by 2π . These are the slope (\dot{A}) and convexity (\ddot{A}) estimators. With the periodogram defined via $I(\lambda) = \sum_{|h| < T} \hat{\gamma}_h e^{-ih\lambda}$ for $\lambda \in [-\pi, \pi]$, it follows that (4) can be re-expressed as $\Xi_{\dot{A}}(I)$ and $\Xi_{\ddot{A}}(I)$, respectively. Two examples of kernels are adapted from the Quartic and Tukey–Hanning kernels discussed in Priestley (1981). The inverse Fourier transforms are given in Appendix F in the supporting information. When calculating (4), it suffices to take the real part of ζ_h in those formulas. The presence of a peak corresponds to $\Xi_A(\dot{f}) = 0$ and $\Xi_A(\ddot{f}) < 0$, because when β is quite small, these conditions correspond to f having a local maximum at $\lambda = \theta$. However, because of the difficulties inherent in adopting a composite null hypothesis corresponding to $\Xi_A(\ddot{f}) < 0$, McElroy & Holan (2009) formulate two simple null hypotheses via

$$\begin{aligned} H_0^{(1)}: \Xi_A(\dot{f}) &= 0 \text{ versus } H_a^{(1)}: \Xi_A(\dot{f}) \neq 0 \\ H_0^{(2)}: \Xi_A(\ddot{f}) &= 0 \text{ versus } H_a^{(2)}: \Xi_A(\ddot{f}) < 0. \end{aligned}$$

Then detection of a peak corresponds to first failing to reject $H_0^{(1)}$, followed by rejecting $H_0^{(2)}$. Under these null hypotheses, the estimators are asymptotically normal when multiplied by \sqrt{T} , with asymptotic variance $\Xi_{\dot{A}^2}(f^2)$ and $\Xi_{\ddot{A}^2}(f^2)$, respectively. These quantities can be consistently estimated by $\sum_{|h|, |k| < T} \hat{\gamma}_h \hat{\gamma}_k \zeta_h - k/2$, where ζ_h is the inverse Fourier transform of either \dot{A}^2 or \ddot{A}^2 . This variance estimate can also be re-expressed as $\Xi_{\dot{A}^2}(I^2)/2$ and $\Xi_{\ddot{A}^2}(I^2)/2$, respectively. Then the test statistics for slope and convexity are

$$\sqrt{T} \frac{\Xi_{\dot{A}}(I)}{\sqrt{\Xi_{\dot{A}^2}(I^2)/2}} \quad \text{and} \quad \sqrt{T} \frac{\Xi_{\ddot{A}}(I)}{\sqrt{\Xi_{\ddot{A}^2}(I^2)/2}}.$$

These are asymptotically standard normal under the null hypotheses, under mild regularity conditions: the fourth-order cumulants of $\{X_t\}$ vanish, and condition B1 of Taniguchi & Kakizawa (2000) holds (this guarantees that the higher-order polyspectra exist and have uniformly continuous derivatives). The size and power properties are adequate (McElroy & Holan, 2009).

The SC procedure involves setting β and θ as desired, and testing slope and convexity in turn, declaring that a peak is present if one fails to reject the slope test but does reject the convexity test (in the lower one-sided direction). If it is desired to test multiple peaks (as in the case of monthly data, where we should consider peaks at the frequencies $\pi j/6$ for $1 \leq j \leq 5$), then we construct multiple kernels by taking different choices of θ and combine the resulting p -values in order to control the family-wise error rate.

2.4 The Visual Significance Diagnostic

In contrast to SC, the VS approach does a direct comparison of the spectral density's value at a frequency θ to two of its close neighbouring frequencies $\theta \pm \delta$. Graphically, this comparison describes two right triangles—adjoining one another—whose angles determine the steepness of the peak. Therefore, it can be deduced that the spacing of the two neighbours relative to the middle ordinate is important and is related to how great the middle value must exceed the neighbouring values to be judged a true peak. This is the basis of the visual significance concept—though the true shape of the peak could be very cusp-like or flat, and we might desire a more nuanced definition. (Appendix A of McElroy & Roy, 2021, provides additional discussion and a parametric class of peaked spectral densities.)

This framework for VS is formalised in McElroy & Roy (2021) through a peak functional defined via

$$\Theta_{\theta, \delta}[g_f] = \min\{g_f(\theta) - g_f(\theta - \delta), g_f(\theta) - g_f(\theta + \delta)\}, \quad (5)$$

where g_f is a suitably chosen monotone transformation of the spectral density f . This is based on the original formulation of Soukup & Findley (1999), which proposed the settings of $g_f = \log f$ and $\delta = \pi/60$. (Another possibility, suggested by a referee, is to take g_f to be a power transformation with power parameter greater than 1—for example, belonging to the family of Box–Cox transformations—to further emphasise the separation of the peak from its immediate neighbourhood.) The VS criterion is that a visually significant peak exists at frequency θ if $\Theta_{\theta, \delta}[\log f]$ exceeds $\tau_f = \tau R_{\log f}$, which depends on the dynamic range of the log spectral density, defined by

$$R_{\log f} = \max_{\lambda \in [-\pi, \pi]} \log f(\lambda) - \min_{\lambda \in [-\pi, \pi]} \log f(\lambda)$$

Various values for τ can be considered; Soukup & Findley (1999) set $\tau = 6/52$ (with $\delta = \pi/60$), but lower values of τ make the VS criterion less demanding.

The VS methodology is rigorously developed in McElroy & Roy (2021), which we review here. The null region for the peak testing problem includes spectral features that resemble half-peaks, where on one side of the target frequency the measure fails to qualify as peaked, even though on the other side the measure may indicate a strongly peaked feature. The right-side and left-side peak measures are defined as $\Theta^{L, \delta}_{\theta}[\log f] = \log f(\theta) - \log f(\theta - \delta)$ and $\Theta^{R, \delta}_{\theta}[\log f] = \log f(\theta) - \log f(\theta + \delta)$, so that $\Theta_{\theta, \delta}[\log f] = \min\{\Theta^{L, \delta}_{\theta}[\log f], \Theta^{R, \delta}_{\theta}[\log f]\}$. Then the null and alternative hypotheses are

$$H^0_{\theta}: \Theta_{\theta, \delta}[\log f] \leq \tau_f \quad \text{versus} \quad H^a_{\theta}: \Theta_{\theta, \delta}[\log f] > \tau_f, \quad (6)$$

which are tested via the test statistic $\Theta_{\theta, \delta}[\log \hat{f}_b]$, where \hat{f}_b is a tapered spectral density estimator of bandwidth bT . This estimator is based upon tapering the first bT sample autocovariances with

a positive definite taper Λ , such as the Bartlett or Daniell kernels (Priestley, 1981), where $b \in (0, 1)$ is a fixed fraction:

$$\hat{f}_b(\lambda) = \sum_h \Lambda(h/(bT)) \hat{\gamma}_h e^{-ih\lambda}.$$

The taper Λ is an even function that typically places more weight on the lower lags, and in many cases, the taper is supported on $[-1, 1]$. The rejection region for the VS test is given by

$$\hat{\mathcal{C}}_\alpha = \{\hat{f}_b : \Theta_{\theta, \delta}[\log \hat{f}_b] > c_\alpha + \tau R_{\log \hat{f}}\},$$

where $R_{\log \hat{f}}$ is the estimated dynamic range of the log spectrum based on an $\text{AR}(p)$ spectral density estimator \hat{f} of f (cf. Tiao & Tsay, 1983). The critical value c_α is the upper $100(1 - \alpha)\%$ point of the distribution of $\Theta_{\theta, \delta}[\log S_\theta(b)]$, where $S_\theta(b)$ is the asymptotic distribution of \hat{f}_b , as given by theorem 2 of McElroy & Roy (2021). This asymptotic theory is based on the fixed- b asymptotics introduced in Hashimzade & Vogelsang (2008); also see McElroy & Politis (2014) and Sun (2014). The asymptotic theory is valid if the process $\{X_t\}$ is linear (with square summable moving average coefficients and finite fourth moments of the inputs) or satisfies the assumptions discussed for the SC procedure. These critical values can be determined by simulation, for each choice of taper and bandwidth fraction b . The size and power properties are generally adequate when the sample size is sufficiently large relative to δ ; at least 8 to 10 years of data is recommended when the Bartlett taper is used.

2.5 The Role of Differencing

A seasonally adjusted series will be integrated of the same order as the original raw data, and hence, seasonally adjusted economic data will typically require differencing to be rendered stationary. This is an important operation for the four seasonal adjustment diagnostics, because all require stationarity as a working assumption. This is because, for the Q_s and ROOT diagnostics, we examine autocovariances—which are only well defined for a stationary process—and for the SC and VS diagnostics, we examine the spectral density, which likewise requires stationarity. Typically, one or two regular differences will be applied to seasonally adjusted data before computing a diagnostic; there is the danger of over-differencing, which introduces non-invertibility into the data. This has various ramifications for each diagnostic, which we consider in turn.

A related issue is that the diagnostics require that there will be no time-varying mean. To see why this is a challenge, consider the case that a time series is not integrated but has a linear time trend; whereas differencing reduces the time trend to a constant (which is not a problem for the diagnostics, because each of them can allow for a constant mean), it will over-difference the stochastic portion of the process, thereby introducing non-invertibility. The correct approach in this case would be to estimate (via regression) the linear time trend and remove it, and avoid differencing the data. In practice, when a linear time trend is present in the data, it can be difficult to discriminate between cases—whether the process is integrated (with a linear time trend implicitly present) or really consists of the sum of a stationary process and a linear time trend. The difficulty of this discernment can produce many instances of over-differencing.

Letting $\{X_t\}$ denote the differenced data process, let us suppose that $X_t = (1 - B)Z_t$, where $\{Z_t\}$ is stationary and invertible; that is, we have over-differenced. Let γ_h^X and γ_h^Z denote the respective autocovariances; these are related via

$$\gamma_h^X = 2\gamma_h^Z - \gamma_{h-1}^Z - \gamma_{h+1}^Z.$$

In the case of Q_s , we are ultimately interested in whether γ_s^Z has a high value, but due to over-differencing, we will be examining γ_s^X , which depends not only on γ_s^Z but also the neighbouring lags γ_{s-1}^Z and γ_{s+1}^Z . Curiously, a type II error can be generated when there is a linear pattern to the autocovariance function γ^Z , such that γ_s^Z is positive but $\gamma_s^X = 0$. Likewise, type I errors can also be generated.

The ROOT diagnostic is based upon fitting a high-order AR model, which will always be mis-specified for a non-invertible process. Empirically, the fit will attempt to approximate the unit moving average root, which will in turn distort the other parameter estimates. In principle, adapting the ROOT diagnostic to allow for non-invertible models (e.g. by fitting ARMA models with the moving average polynomial allowed to have unit roots) would rectify the distortion due to over-differencing, because the model would automatically account for this effect—but as currently implemented, ROOT will be adversely impacted by over-differencing.

The impact of over-differencing in the frequency domain is to introduce a high-pass filter to the data: the spectral densities f^X and f^Z are related via

$$f^X(\lambda) = (2 - 2\cos(\lambda))f^Z(\lambda).$$

So the action of the high-pass filter function $2 - 2\cos(\lambda)$ eradicates any frequency content near $\lambda = 0$ and greatly attenuates features in f^Z in a neighbourhood of the origin. Hence, the first seasonal peak at $2\pi/s$ in f^Z can be attenuated and/or shifted by the high-pass filter, thereby interfering with its identification through statistics based on f^X . Hence, VS and SC may be adversely impacted, potentially having less capability to identify seasonal peaks, thereby generating type I error.

3 Data Analysis

3.1 Empirical Analysis of Series from Multiple Sectors

We explore 233 monthly time series published by the U.S. Census Bureau (see <https://www.census.gov/retail/index.html>), composed of 65 Retail Trade and Food Services (MRTS) time series, 22 Wholesale Trade: Sales and Inventories (MWTS) time series, 4 Manufacturers' Shipments, Inventories, and Orders (M3) time series, 87 Manufacturing and Trade Inventories and Sales (MTIS) time series and 54 New Residential Construction (RES) time series. The start dates for MRTS, MWTS, M3 and MTIS series are January 1992 or later and that for RES is January 1959 or later. The end date for all series is September 2019. Each series displays some degree of seasonal persistence, as well as trend dynamics, extremes and holiday effects. We seasonally adjust each series via the X-13ARIMA-SEATS (U.S. Census Bureau, 2015) software, via either the X11 option (with non-parametric filters applied to ARIMA forecast extensions) or the SEATS option (with parametric filters).

The automatic modelling option of the software is used, so that calendrical effects and extremes are accounted for; differencing polynomials and log transformations are also identified. To each adjustment, we apply the SC, VS, Q_s and ROOT diagnostics. To remove the edge effect of the seasonal adjustment, we trim the first and the last 3 years of data before applying the diagnostics.

The multiple seasonal frequency version of the ROOT diagnostics is used for the five seasonal frequencies with a ρ_0 value of 0.97 corresponding to each of the roots. For the VS diagnostic, the values of the different tuning parameters are $\tau = 0.1$ and $b = 0.6$. The VS and the SC diagnostics are frequency-domain tests; the p -values for tests of inadequacy of adjustment

(nominal level 5%) at the five seasonal frequencies are combined using the Hochberg (1988) procedure for controlling the family-wise error rate (FWER). The FWER control level is $\alpha = 0.05$, and the seasonal adjustment is deemed inadequate if the adequacy hypothesis is rejected for at least one seasonal frequency.

For each of the 232 original series, the tests are applied to four post-adjustment series, namely, the seasonally adjusted series (for both the X11 and SEATS options) and the irregular components under the two options. The VS diagnostic did not flag any series for inadequate adjustment, but the other diagnostics did identify a few problematic series. Because there is no specific pattern in terms of agreement or disagreement between the diagnostics, the full number of series flagged is presented in Table 2.

An overwhelming portion of the series are deemed to be adequately adjusted by all the diagnostics. In fact, VS does not identify any problems; in contrast, SC identifies a total of 27 instances of inadequate adjustment (when we tally over the SA and Irr components for both the X11 and SEATS options). The ROOT and Q_s diagnostics flag 7 and 10 instances, respectively. There were no discernible patterns in the way the different diagnostics deemed inadequate adjustment. The disparate results among the diagnostics are not unexpected, because these diagnostics have different null hypotheses (i.e. different mathematical formulations of seasonality) and have different power profiles.

3.2 Diagnosis of Residual Seasonality for Problem Series

In the preceding section’s study, not knowing the true status of adjustments (whether there truly is residual seasonality present) means it is difficult to compare the performance of the different diagnostics. Each diagnostic will have false positive and false negative determinations. Ideally, one would devise a simulation experiment to compare performances, but given that the null hypothesis associated with different diagnostics permits different types (and degrees) of seasonal features, it is difficult to envision a proper simulation setting where one could make an apples to apples comparison. (A study of performance on synthetic data is given in the next subsection.) Here, we undertake a different approach to testing, whereby an inadequate adjustment is intentionally generated (using a filter different from that determined by the seasonal adjustment software), and tally for how many such cases each diagnostic is able to (correctly) detect residual seasonality.

In the succeeding text, we examine in more detail 15 time series that—on the basis of prior analysis described in Findley *et al.* (2017)—are known to have problematic seasonal adjustments. In particular, we study 15 time series of the U.S. Census Bureau Monthly Retail Trade Survey that are suspected to exhibit changing seasonality (based on the fact that X-13ARIMA-

Table 2. Number of series seasonally adjusted by X11 or SEATS that are deemed to be adequate, according to Q_s test, the ROOT test or the SC test, for either the seasonally adjusted (SA) component or the irregular (Irr).

| Test | X11 | | | | | SEATS | | | | |
|-----------|------|------|----|------|-----|-------|------|----|------|-----|
| | MRTS | MWTS | M3 | MTIS | RES | MRTS | MWTS | M3 | MTIS | RES |
| SA Q_s | 64 | 22 | 87 | 4 | 52 | 64 | 22 | 86 | 4 | 51 |
| SA ROOT | 63 | 22 | 87 | 4 | 54 | 65 | 22 | 86 | 4 | 54 |
| SA VS | 65 | 22 | 87 | 4 | 54 | 65 | 22 | 87 | 4 | 54 |
| SA SC | 63 | 22 | 86 | 4 | 54 | 65 | 22 | 86 | 4 | 47 |
| Irr Q_s | 63 | 22 | 87 | 4 | 54 | 65 | 22 | 87 | 4 | 54 |
| Irr ROOT | 64 | 22 | 87 | 4 | 54 | 65 | 22 | 85 | 4 | 53 |
| Irr VS | 65 | 22 | 87 | 4 | 54 | 65 | 22 | 87 | 4 | 54 |
| Irr SC | 63 | 21 | 84 | 4 | 51 | 65 | 22 | 85 | 4 | 47 |
| Total | 65 | 22 | 87 | 4 | 54 | 65 | 22 | 87 | 4 | 54 |

SEATS automatically selected short filters for each of the series); hence, each calendar month's average seasonality for a period of 8 years is expected to differ from that for the entire 16-year span. The series titles are given in Table 3, with start and end dates of January 1992 and December 2007, respectively.

For each series, we perform a 'bad' seasonal adjustment by simply removing monthly averages computed over 16 years. Because of our prior evidence that these series have rapidly changing seasonality, we expect that such a stable seasonal adjustment—based on supposing that the seasonality is rigid and unchanging—will be inadequate. We apply the four diagnostics to the last 8 years (of the trimmed data, where we remove the first and last 2 years) of such adjustments and report performances in Table 3. For the VS diagnostic, the width parameter is chosen to be $\pi/15$ due to the relatively shorter length of the series. All other settings are the same as that of the previous subsection.

For seven of the 15 series, none of the diagnostics detect any residual seasonality. Overall, Q_s is the most sensitive, which is unsurprising given that any degree of seasonal autocorrelation will asymptotically cause the null hypothesis (of no residual seasonality) to be rejected. The ROOT and the VS diagnostic identify the same set of series, and the set is a subset of those flagged by Q_s . Curiously, while the SC flags only two series, one of them, 45100, is not identified by the other three diagnostics. Upon closer examination, we find that spectral peaks are sharper over the smaller span, as opposed to the whole data, whereas the sample autocorrelation (for the differenced series) at lags 12 and 24 is quite small. So Q_s , which is based on autocorrelation, fails to detect moving seasonality. The second spectral peak is quite broad, and perhaps this specific form of seasonality triggers the SC diagnostic and not the other tests.

3.3 Evaluation on synthetic series

To study the performance of the diagnostics on simulated time series that have characteristics resembling real data, we create synthetic series by taking a convex combination of the seasonal factor and seasonal adjustment of real series. In particular, for any real series, we obtain its seasonal adjustment and seasonal factor (using the X11 option of X-13ARIMA-SEATS) and construct a new synthetic series by adding a scalar multiple λ of the seasonal factor to the seasonal adjustment. Hence, if $\lambda = 0$, there is no seasonality present (presuming the output seasonal

Table 3. Diagnostic indications of residual seasonality in the 8-year period (January 1998 through December 2005) of the stable seasonal adjustment, for 15 series of the U.S. Census Bureau Monthly Retail Trade Survey.

| Title | Series | Q_s | ROOT | VS | SC |
|--|--------|-------|------|-----|----|
| Retail and food services sales, total | 44000 | - | - | - | - |
| Electronics and appliance stores | 44300 | *** | *** | *** | * |
| Computer and software stores | 44312 | ** | - | - | - |
| Building materials and supplies dealers | 44400 | - | - | - | - |
| Grocery stores | 44510 | - | - | - | - |
| Clothing and clothing accessory stores | 44800 | - | - | - | - |
| Women's clothing stores | 44812 | ** | - | - | - |
| Shoe stores | 44820 | ** | - | - | - |
| Sporting goods, hobby, book and music stores | 45100 | - | - | - | * |
| General merchandise stores | 45200 | *** | *** | ** | - |
| Department stores—excluding leased departments | 45210 | - | - | - | - |
| Warehouse clubs and superstores | 45291 | *** | *** | *** | - |
| Non-store retailers | 45400 | - | - | - | - |
| Electronic shopping and mail-order houses | 45410 | *** | *** | ** | - |
| Food services and drinking places | 72200 | - | - | - | - |

The type I error/FWER levels of significance are *** for <0.001 , ** for <0.05 , * for <0.1 and - for ≥ 0.1 .

adjustment was adequate), whereas $\lambda = 1$ corresponds to reconstructing the original series. (This recipe holds for additive seasonal adjustments; in the case of a multiplicative seasonal adjustment, we apply the earlier prescription to the log of the seasonal adjustment and seasonal factor.) We can even let $\lambda > 1$, if we want to make the seasonality more prevalent than it was in the original series.

This construction generates a varying degree of seasonality in a synthetic series but based on the features present in real data rather than through a simulation. This has the advantage of giving a more fair basis of comparison of the diagnostics, without giving any unfair advantage due to particular methodological assumptions being satisfied. We can make the construction more exotic by introducing seasonal heteroscedasticity, as follows: for one of the months, we scale the seasonal factor by 5λ , whereas the other 11 months are scaled by λ . This generates more seasonal variability in that particular month (which we take to be January). Note that seasonal heteroscedasticity violates the working assumptions of all four diagnostics, which presume that the appropriately differenced (and windowed) process is stationary; season-dependent variability violates the stationarity assumption. Nevertheless, we may still hope that the diagnostics are able to detect some seasonality even in such a case.

We apply this synthetic construction to the 15 time series described in Table 3. The settings for the four diagnostics are the same as given in Section 3.1, namely, $\alpha = 0.05$ for all tests, $\tau = 0.1$ and $b = 0.6$ for VS, and $\rho_0 = 0.97$ for ROOT. Taking increments of λ of size 0.01, and beginning with 0, we increase λ up to 5 and record the smallest value λ_\star such that seasonality is detected. If we take the perspective that any non-zero value of λ corresponds to a seasonal synthetic series, then diagnostics with lower values of λ_\star indicate the test is more sensitive to the detection of seasonality, as compared with diagnostics with a higher value of λ_\star . Results are provided in Table 4.

This study indicates that QS, followed by ROOT, tends to be the most sensitive to the presence of seasonality; this makes sense given that any degree of non-zero autocorrelation at seasonal lags is accorded seasonal status by QS. In contrast, VS and SC are more demanding in their criteria, requiring higher values of λ_\star in general to identify the presence of seasonality. (For some series, though, SC is more sensitive than VS.) Of course, these patterns of identification depend on the settings of the individual tests, through the choice of τ , ρ_0 and so forth.

Table 4. Value of λ_\star such that seasonality is detected for that synthetic series for any $\lambda \geq \lambda_\star$ and is not detected for $\lambda < \lambda_\star$.

| Series | Homoscedastic | | | | Heteroscedastic | | | |
|--------|---------------|------|------|------|-----------------|------|------|------|
| | Qs | ROOT | VS | SC | Qs | ROOT | VS | SC |
| 44000 | 0.66 | 0.91 | 1.04 | 2.55 | 0.27 | 0.40 | 0.59 | 1.25 |
| 44300 | 0.33 | 0.27 | 1.87 | 1.16 | 0.27 | 0.27 | — | 0.82 |
| 44312 | 0.42 | 0.73 | 2.64 | 2.41 | 0.29 | 0.51 | — | 1.50 |
| 44400 | 0.82 | 1.10 | 1.34 | 1.91 | 0.85 | 1.11 | 1.27 | 0.65 |
| 44510 | 0.74 | 0.90 | 3.53 | 3.38 | 0.30 | 0.43 | 1.14 | 0.54 |
| 44800 | 0.77 | 0.64 | 0.54 | 2.52 | 0.29 | 0.37 | 0.89 | 0.73 |
| 44812 | 0.77 | 0.65 | 1.08 | 2.46 | 0.26 | 0.36 | 0.63 | 0.80 |
| 44820 | 0.61 | 0.73 | 3.41 | 2.73 | 0.25 | 0.29 | 0.88 | 0.51 |
| 45100 | 0.57 | 0.70 | 3.25 | 2.60 | 0.27 | 0.46 | 1.09 | 0.96 |
| 45200 | 0.36 | 0.39 | 0.96 | 1.03 | 0.15 | 0.19 | 1.00 | 0.59 |
| 45210 | 0.68 | 0.61 | 1.96 | 2.40 | 0.20 | 0.24 | 1.07 | 0.77 |
| 45291 | 0.31 | 0.40 | 0.73 | 0.94 | 0.17 | 0.26 | 0.42 | 0.39 |
| 45400 | 0.59 | 0.81 | 0.74 | 2.24 | 0.36 | 0.47 | 0.68 | 0.67 |
| 45410 | 0.51 | 0.43 | 0.65 | 2.04 | 0.34 | 0.35 | 0.66 | 1.13 |
| 72200 | 1.26 | 1.31 | 2.37 | 1.92 | 0.39 | 0.64 | 1.74 | 1.23 |

The blank entries for VS indicate that no such $\lambda_\star \leq 5$ was obtained.

We also find that the four diagnostics are able to identify heteroscedastic seasonality as well, even though they are not explicitly designed to do so, and generally, a lower value of λ_* is obtained as compared with the homoscedastic case. However, this may simply be due to the January month having five times as much seasonal amplitude as the other months do in the homoscedastic scenario.

4 Discussion

One of the main tasks for statistical offices is publishing seasonally adjusted time series; therefore, determining whether adjustments are adequate is critically important. While this task requires training and sober judgement in normal times, the job has become more challenging during the Covid-19 pandemic. Starting in spring 2020, the time series helpdesk at the U.S. Census Bureau received inquiries (from the NYC Mayor's office, and the National Accounts of Peru and Columbia, etc.) about modelling time series affected by the pandemic. Internally, regular working groups met to discuss the use of outliers to model extreme values, with a view towards maintaining the quality of published seasonal adjustments. Although the same sorts of diagnostic tools can be used during a time of crisis, they may be used differently, and new questions arise—because extreme values are no longer isolated, as in more regular epochs, but may be clustered into strange patterns. One outcome has been the recognition of the need for more research into ascertaining when there is a return to normalcy from an epoch where exotic kinds of extremes are prevalent.

Whether or not a crisis is in play, the determination of whether a series is a candidate for seasonal adjustment in the first place is an important topic; however, this determination requires diverse tools from the problem of assessing adequacy of seasonal adjustment and is essentially a different problem. Because unit root tests are not appropriate for diagnosing seasonal adjustment adequacy, this review focuses on four tests (Q_s , ROOT, VS and SC) that are based upon time-domain and frequency-domain conceptions of stationary seasonality. Q_s is flawed but is given an updated asymptotic theory (Appendix B in the supporting information) that somewhat improves performance; in contrast, the ROOT diagnostic is a flexible tool showing promise. VS has a more nuanced conception of peak than that espoused by the SC procedure. These statistics are tested in simulation (Appendices C and E for Q_s and ROOT; SC and VS have been assessed in McElroy & Holan (2009) and McElroy & Roy (2021), respectively) and applied to a large collection of monthly time series (as well as synthetic series), for which we offer a comparison of the four methods.

We do not endorse the Q_s and SC methods, given that the ROOT and VS methods are available and offer better performance and discrimination. In particular, Q_s can give spurious indications of seasonality even when the correct asymptotic critical values are used. The ROOT diagnostic avoids this problem by focusing on the oscillatory character of autocorrelations, and the persistency of the seasonal effect can be incorporated—allowing one to discriminate between mild and strong forms of stationary seasonality. Similarly, the VS diagnostic in the frequency domain can discriminate between different degrees of seasonality based on the steepness of the spectral peak; however, it is crucial to use the correct asymptotic critical values, with the requisite sample size depending upon the specified width of spectral peak. Because SC cannot specify a degree of seasonality in the null hypothesis, it is not as flexible as VS.

What are the practical ramifications of seasonal adjustment? Seasonal adjustment inevitably modifies a time series in unintended ways. If the seasonality is deterministic, then regression effects can be estimated and subtracted, causing minor distortions through the parameter estimates. Distortions can be much greater when using filters—for suppressing dynamic seasonality—because the filters may depend on fitted model parameters and are typically

applied to data that have been forecast extended. These extensions generate local non-linearity at the boundaries of the sample (Dagum, 1980); non-linear edge effects can be isolated by truncating a few years of data from the beginning and end of the time series (Findley *et al.*, 1998). Even in an ideal case, in the centre of the sample, many model-based filters achieve optimal seasonal adjustment by over-suppressing seasonal patterns, thereby generating negative autocorrelations at seasonal lags; this is the over-adjustment problem. In addition to these distortions, extreme values can generate challenges for seasonal adjustment diagnostics; the asymptotic theory that we have discussed requires finite fourth moments and a low degree of kurtosis.

Diagnostics must be adapted to the characteristics of seasonally adjusted data; the currently available diagnostics (Q_s , ROOT, SC and VS) are oriented towards the problem of under-adjustment, being designed to test whether any seasonality remains. These diagnostics can be modified to address the secondary issue of over-adjustment as well. However, there are other distortions (e.g. non-linear and locally non-stationary effects) that seasonal adjustment generates, and typically, there is little concern about this evinced in the seasonal adjustment community. W. R. Bell proposed the use of concurrent filters for all seasonal adjusting [which the decompositions of Brewer (1979) and Proietti (1995) achieve], hence removing the forecast extensions and thereby eliminating introduced local non-stationarity. If the published seasonal adjustment is viewed as a final product—that is not needed for other subsequent statistical analyses—then the use of symmetric filters is preferable to concurrent filters, because of increased accuracy. However, the lesser accuracy resulting from the use of concurrent filters may be acceptable in view of the fact that seasonally adjusted data are commonly used in other applications, such as forecasting—in this connection, it is preferable to introduce as few distortions to the time series structure as possible. In any case, there is an opportunity for additional research to devise a wider array of diagnostic tools that would examine more than just the issue of under-adjustment.

Supposing now that non-Gaussian and non-linear structures have been accounted for in a time series, and a seasonal adjustment has been flagged as inadequate by one or more diagnostics, then the analyst will have to readjust the series. Now, the nature of the different diagnostics can be useful: VS and SC indicate bad seasonal adjustment if there are residual seasonal peaks in the spectral density, whereas Q_s and ROOT flag problems with the autocorrelations. Usually, the problem can be resolved by altering the filters such that smoothing is more localised, corresponding to a wider trough in the frequency response function of the seasonal adjustment filter (McElroy, 2012). Sometimes, the fitted model—either used for forecast extension or to determine model-based filter weights—needs to be modified and improved, and altering the data span by omitting some earlier years quite often is a sufficient remedy. (It is not really necessary to obtain a model that reduces the data to white noise residuals, but only an end result of seasonal adjustment adequacy is needful.) Conversely, over-adjustment can be rectified by broadening the filter bandwidth, corresponding to a narrower trough in the seasonal adjustment filter's frequency response function. Very rarely will it be necessary to alter the seasonal adjustment method. This could occur in a case where non-parametric filters (such as those of X-12-ARIMA) are being used, but there is insufficient variety in the filter class to properly remove the seasonality. Switching to a model-based framework (such as implemented in SEATS) may allow for a richer class of filters, increasing the likelihood that a particular filter that is well adapted to the data will be utilised.

In looking forward to the future of seasonal adjustment, there are many avenues of further research. There is a wide scope for devising new measures of residual seasonality. The discussion in this paper has focused upon stationary processes for which the autocorrelation function and spectral density are well defined, but heteroscedastic processes can exhibit a form of non-stationarity that is seasonal, exhibiting larger variation (or levels) in particular seasons.

One approach is to use the stacking idea of Tiao & Grupe (1980) to obtain a stationary multivariate time series of lower frequency and assess the dynamics of the process. In this connection, Lin *et al.* (2020) develop a non-parametric model for seasonal dynamics through a singular value decomposition of the data matrix—but this technique has not been developed into a tool for measuring seasonality. Another possibility is to consider time-varying estimates of the spectral density, which may be helpful when local non-stationarity is present; possibly, manifestations of under-adjustment or over-adjustment might only be present in subspans of a time series, requiring a localised frequency-domain approach—for example, by complex demodulation (Hasan, 1983).

Within the class of stationary processes, there seems to be some promise with developing a diagnostic based upon inverse partial autocorrelations (discussed in Appendix A in the supporting information). Moving to frequency domain, there may be value in developing diagnostics based upon polyspectra, such as the bispectrum. A polyspectral diagnostic could be useful for non-linear processes, because it is possible to have peaks (say, in the bispectrum) even when the spectral density is flat. The repercussions of such higher-order seasonality have not been developed; all current treatment is based on first and second moment structures of the time series. The appeal of such a framework for investigating higher-order seasonality is that this approach is not tied to a specific class of non-linear processes or models and hence can be equally applied to GARCH, non-Gaussian SARMA or more exotic (but stationary) processes; a useful diagnostic should not be embedded within a narrow framework, because the range of data requiring seasonal adjustment is remarkably diverse.

5 Disclaimer

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

Data Availability Statement

Data, as well as R codes for all procedures, are available upon request.

Notes

¹The term ‘persistence’ is used to denote a high degree of association in a stochastic process between various random variables, possibly measured through autocorrelation, spectra or fixed effects (regressors). The exact definition will depend on whether the stochastic process is stationary or non-stationary, and the form the persistence takes.

²By including in the seasonal adjustment filter a polynomial in the lag operator B of the form $1 + B + \dots + B^{s-1}$, where s is the seasonal period, all period s functions will be annihilated by the filter, and any seasonal unit root process will be rendered stationary; see remark 3.5.9 of McElroy & Politis (2020).

³Based on the analysis of the SAR process in Findley *et al.* (2015), a conservative choice is $\tau = 0.7$, because lower values of seasonal autocorrelation imply a negligible degree of autocorrelation after 4 years.

REFERENCES

Akaike, H. 1980. Seasonal adjustment by a Bayesian modeling. *J. Time Series Anal.*, **1**(1), 1–13.

International Statistical Review (2022), **90**, 2, 259–284

Published 2021. This article is a U.S. Government work and is in the public domain in the USA.

- Akaike, H. & Ishiguro, M. 1983. Comparative study of the X-11 and BAYSEA procedures of seasonal adjustment. *Applied Time Series Analyses of Economic Data*.
- Ansley, C. & Wecker, W. 1984. Issues involved with the seasonal adjustment of economic time series, comment: On dips in the spectrum of a seasonally adjusted time series. *J. Bus. Econ. Stat.*, **2**(4), 323–324. <https://doi.org/10.1080/07350015.1984.10509400>
- Bell, W.R. 2004. On RegComponent time series models and their applications. In *State Space and Unobserved Component Models: Theory and Applications (Proceedings of a Conference in Honour of James Durbin)*, Eds. A. Harvey, S. J. Koopman & N. Shephard, Cambridge University Press: Cambridge, pp. 248–283.
- Bell, W.R. & Hillmer, S.C. 1983. Modeling time series with calendar variation. *J. Am. Stat. Assoc.*, **78**, 526–534.
- Bell, W.R. & Hillmer, S.C. 1984. Issues involved with the seasonal adjustment of economic time series. (Invited paper with discussion). *J. Bus. Econ. Stat.*, **2**, 299–320.
- Bell, W.R. & Hillmer, S.C. 1985. A reply (to comments by Sims in Bell and Hillmer (1984)). *J. Bus. Econ. Stat.*, **3**, 95–97.
- Blakely, C. & McElroy, T.S. 2017. Signal extraction goodness-of-fit diagnostic tests under model parameter uncertainty: Formulations and empirical evaluation. *Econ. Rev.*, **36**(4), 447–467. <https://doi.org/10.1080/07474938.2016.1140277>
- Brewer, K. 1979. Seasonal adjustment of ARIMA series. *Economie appliquée Paris*, **32**(1), 7–22.
- Brewer, K., Hagan, P. & Perazzelli, P. 1975. Seasonal adjustment using Box-Jenkins models. In *Bulletin of the International Statistical Institute Proceedings of the 40th Session*, **31**, 130–136.
- Brockwell, P.J. & Davis, R.A. 2013. *Time Series: Theory and Methods*. Springer Science & Business Media.
- Busetti, F. & Harvey, A. 2003. Seasonality tests. *J. Bus. Econ. Stat.*, **21**(3), 420–436.
- Buys Ballot, C.H.D. 1847. *Les changements périodiques de température*. Keminik et Fils: Utrecht.
- Canova, F. & Hansen, B.E. 1995. Are seasonal patterns constant over time? A test for seasonal stability. *J. Bus. Econ. Stat.*, **13**(3), 237–252.
- Chen, R. & Fomby, T.B. 1999. Forecasting with stable seasonal pattern models with an application to Hawaiian tourism data. *J. Bus. Econ. Stat.*, **17**(4), 497–504.
- Cleveland, W.S. & Devlin, S.J. 1980. Calendar effects in monthly time series: Detection by spectrum analysis and graphical methods. *J. Am. Stat. Assoc.*, **75**(371), 487–496.
- Cleveland, W.S. & Terpenning, I.J. 1982. Graphical methods for seasonal adjustment. *J. Am. Stat. Assoc.*, **77**(377), 52–62.
- Dagum, E.B. 1980. The X-11-ARIMA seasonal adjustment method. Statistics Canada, <https://www.census.gov/ts/papers/1980X11ARIMAManual.pdf> (accessed June 3, 2016).
- del Barrio Castro, T. & Osborn, D.R. 2004. The consequences of seasonal adjustment for periodic autoregressive processes. *The Econ. J.*, **7**(2), 307–321.
- den Butter, F. & Fase, M. 1991. *Seasonal Adjustment as a Practical Problem*. Amsterdam: North Holland.
- Fase, M.M., Konig, J. & Volgenants, A.F. 1973. An experimental look at seasonal adjustment. *De Economist*, **121**, 441–486.
- Findley, D.F. 2005. Some recent developments and directions in seasonal adjustment. *J. Off. Stat.*, **21**, 343–365.
- Findley, D.F., Lytras, D.P. & Maravall, A. 2015. Illuminating model-based seasonal adjustment with the first order seasonal autoregressive and airline models. Census Bureau Research Report: RRS2015/02, <https://www.census.gov/srd/papers/pdf/RRS2015-02.pdf>
- Findley, D.F., Lytras, D.P. & McElroy, T.S. 2017. Detecting seasonality in seasonally adjusted monthly time series. Census Bureau Research Report: RRS2017/03, <https://www.census.gov/srd/papers/pdf/RRS2017-03.pdf>
- Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. & Chen, B.C. 1998. New capabilities of the X-12-ARIMA seasonal adjustment program (with discussion). *J. Bus. Econ. Stat.*, **16**, 127–177. <https://www.census.gov/ts/papers/jbes98.pdf>
- Franses, P.H. 1994. A multivariate approach to modeling univariate seasonal time series. *J. Econ.*, **63**(1), 133–151.
- Franses, P.H. 1996. Recent advances in modelling seasonality. *J. Econ. Surv.*, **10**(3), 299–345.
- Franses, P.H. & Paap, R. 2004. *Periodic Time Series Models*. OUP: Oxford.
- Friedman, M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.*, **32**(200), 675–701.
- Ghysels, E. & Osborn, D.R. 2001. *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press: Cambridge.
- Gladyshev, E.G. 1961. Periodically correlated random sequences. *J. Sov. Math.*, **2**, 385–388.
- Granger, C. 1978. *Seasonality: Causation, interpretation, and implications* Edited by A. Zellner, Seasonal Analysis of Economic Time Series, pp. 33–46. NBER: Cambridge, MA.
- Gregoir, S. 2006. Efficient tests for the presence of a pair of complex conjugate unit roots in real time series. *J. Econ.*, **130**(1), 45–100.

- Grether, D.M. & Nerlove, M. 1970. Some properties of “optimal” seasonal adjustment. *Econometrica*, **38**, 682–703.
- Hasan, T. 1983. Complex demodulation: Some theory and applications. *Handbook Stat.*, **3**, 125–156.
- Hashimzade, N. & Vogelsang, T. 2008. Fixed-b asymptotic approximation of the sampling behaviour of nonparametric spectral density estimators. *J. Time Ser. Anal.*, **29**, 142–162.
- Hochberg, Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**(4), 800–802.
- Hylleberg, S. 1986. *Seasonality in Regression*. Academic Press: Orlando, FL.
- Hylleberg, S., Engle, R.F., Granger, C.W. & Yoo, B.S. 1990. Seasonal integration and cointegration. *J. Econ.*, **44**(1–2), 215–238.
- Kruskal, W.H. & Wallis, W.A. 1952. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.*, **47**(260), 583–621.
- Ladiray, D. 2012. Theoretical and real trading-day frequencies. In *Economic Time Series: Modeling and Seasonality*, Chapman and Hall: New York, pp. 255–279.
- Ladiray, D. & Quenneville, B. 2012. *Seasonal Adjustment with the X-11 Method*, Vol. **158**. Springer Science & Business Media.
- Lin, W., Huang, J. & McElroy, T. 2020. Time series seasonal adjustment using regularized singular value decomposition. *J. Bus. Econ. Stat.*, **38**(3), 487–501.
- Lytras, D.P., Feldpausch, R.M. & Bell, W.R. 2007. Determining seasonality: A comparison of diagnostics from X-13-ARIMA. In *Proceedings of the Third International Conference on Establishment Surveys*. <https://www.census.gov/srd/www/sapaper/sapaper.html>
- Maravall, A. 2003. A class of diagnostics in the ARIMA-model-based decomposition of a time series. In *Seasonal Adjustment*, Eds. M. Manna & R. Peronaci: Frankfurt am Main, pp. 23–36.
- Maravall, A. 2012. Update of seasonality tests and automatic model identification in TRAMO-SEATS. Bank of Spain (November 2012 draft).
- Marple, S.L. 1987. *Digital spectral analysis with applications*, Vol. **5**. Prentice Hall: Englewood Cliffs, NJ.
- McElroy, T.S. 2008. Statistical properties of model-based signal extraction diagnostic tests. *Commun. Stat. Theory Methods*, **37**, 591–616. <https://doi.org/10.1080/03610920701669785>
- McElroy, T. 2012. An alternative model-based seasonal adjustment that reduces residual seasonal autocorrelation. *Taiwan Econ. Forecast Policy*, **43**, 33–70.
- McElroy, T. 2018. Seasonal adjustment subject to accounting constraints. *Statistica Neerlandica*, **72**, 574–589.
- McElroy, T. 2021. A diagnostic for seasonality based upon polynomial roots of ARMA models. *J. Off. Stat.*, **37**(2), 1–28.
- McElroy, T. & Holan, S. 2009. A nonparametric test for residual seasonality. *Survey Meth.*, **35**, 67–83.
- McElroy, T. & Politis, D. 2014. Spectral density and spectral distribution inference for long memory time series via fixed-b asymptotics. *J. Econ.*, **182**, 211–225.
- McElroy, T.S. & Politis, D. 2020. *Time Series: A First Course with Bootstrap Starter*. Chapman and Hall: New York.
- McElroy, T. & Roy, A. 2021. Testing for adequacy of seasonal adjustment in the frequency domain. *J. Stat. Plan. Infer.*, **221**, 241–255.
- McNulty, M.S. & Huffman, W.E. 1989. The sample spectrum of time series with trading day variation. *Econ. Lett.*, **31**(4), 367–370.
- Nerlove, M. 1964. Spectral analysis of seasonal adjustment procedures. *Econometrica*, **32**, 241–286.
- Newton, H.J. & Pagano, M. 1983. A method for determining periods in time series. *J. Am. Stat. Assoc.*, **78**(381), 152–157.
- Osborn, D.R. 1988. Seasonality and habit persistence in a life cycle model of consumption. *J. Appl. Econ.*, **3**(4), 255–266.
- Osborn, D.R. 1991. The implications of periodically varying coefficients for seasonal time-series processes. *J. Econ.*, **48**(3), 373–384.
- Piccolo, D. 1982. A comparison of some alternative decomposition methods for ARMA models. *Time Series Anal.: Theory Practice*, **1**, 565–582.
- Pierce, D.A. 1976. Uncertainty in seasonal adjustment procedures. In *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pp. 528–533. Washington, D.C.
- Pierce, D.A. 1979. *Seasonal adjustment when both deterministic and stochastic seasonality are present* Edited by A. Zellner, Seasonal Analysis of Economic Time Series, pp. 242–280. NBER.
- Priestley, M.B. 1981. *Spectral Analysis and Time Series*. Academic Press: New York.
- Proietti, T. 1995. The Beveridge-Nelson decomposition: Properties and extensions. *J. Italian Stat. Soc.*, **4**(1), 101–124.
- Proietti, T. 1996. Persistence of shocks on seasonal processes. *J. Appl. Econ.*, **11**(4), 383–398.
- Proietti, T. 2000. Comparing seasonal components for structural time series models. *Int. J. Forecast.*, **16**(2), 247–260.

- Sims, C. 1978. *Comments on 'Seasonality: Causation, interpretation, and implications' by Clive W.J. Granger Edited by Zellner, A., Seasonal Analysis of Economic Time Series*, pp. 47–49. NBER: Cambridge, MA.
- Soukup, R.J. & 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. In *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 144–149.
- Soukup, R.J. & Findley, D.F. 2001. Detection and modeling of trading day effects. In *ICES II: Proceedings of the Second International Conference on Establishment Surveys*, pp. 743–753.
- Sun, Y. 2014. Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *J. Econ.*, **178**, 659–677.
- Taniguchi, M. & Kakizawa, Y. 2000. *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag: New York City, NY.
- Tiao, G. & Tsay, R. 1983. Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Ann. Stat.*, **11**, 856–871.
- Tiao, G.C. & Grupe, M.R. 1980. Hidden periodic autoregressive-moving average models in time series data. *Biometrika*, **67**, 365–373.
- Tukey, J. 1978. *Comments on 'Seasonality: Causation, interpretation, and implications' by Clive W.J. Granger Edited by Zellner, A., Seasonal Analysis of Economic Time Series*, pp. 50–54. NBER: Cambridge, MA.
- U.S. Census Bureau 2015. X-13-ARIMA Reference Manual. U.S. Census Bureau, Washington, DC, USA. <https://www.census.gov/ts/x13as/docX13AS.pdf>

[Received January 2021; accepted October 2021]