Journal of Official Statistics, Vol. 37, No. 2, 2021, pp. 367–394, http://dx.doi.org/10.2478/JOS-2021-0016

A Diagnostic for Seasonality Based Upon Polynomial Roots of ARMA Models

Tucker McElroy[1]

Methodology for seasonality diagnostics is extremely important for statistical agencies, because such tools are necessary for making decisions whether to seasonally adjust a given series, and whether such an adjustment is adequate. This methodology must be statistical, in order to furnish quantification of Type I and II errors, and also to provide understanding about the requisite assumptions. We connect the concept of seasonality to a mathematical definition regarding the oscillatory character of the moving average (MA) representation coefficients, and define a new seasonality diagnostic based on autoregressive (AR) roots. The diagnostic is able to assess different forms of seasonality: dynamic versus stable, of arbitrary seasonal periods, for both raw data and seasonally adjusted data. An extension of the AR diagnostic to an MA diagnostic allows for the detection of over-adjustment. Joint asymptotic results are provided for the diagnostics as they are applied to multiple seasonal frequencies, allowing for a global test of seasonality. We illustrate the method through simulation studies and several empirical examples.

Key words: Autoregressive estimator; seasonal adjustment; spectral peaks; visual significance.

1. Introduction

The problem of identifying seasonality in published time series is of enduring importance. Many official time series – such as gross domestic product (GDP) data – have an enormous impact on public policy, and are heavily scrutinized by economists and journalists. Obscuring the issue is the lack of universally agreed-upon criteria for detecting seasonality. Furthermore, the tools that critics use to assess seasonality (e.g., seasonal averages of growth rates, as in Rudebusch et al. 2015) sometimes differ from the diagnostics actually employed at statistical agencies, such as Visual Significance (Soukup and Findley 1999; McElroy and Roy 2017), the Q_s diagnostic of Maravall (Findley et al. 2017), and the model-based F test (Lytras et al. 2007).

An overview of seasonality diagnostics that are currently available in popular statistical software is given in Findley et al. (2017); also see discussions in Fase et al. (1973) and Den Butter and Fase (1991). Desiderata for seasonality diagnostics include:

1. a rigorous statistical theory,
2. a precise correspondence between actual seasonal dynamics and diagnostic values,

[1] Research and Methodology Directorate, U.S. Census Bureau, 4600 Silver Hill Road,Washington, D.C. 20233-9100, U.S.A. Email: tucker.s.mcelroy@census.gov
Acknowledgments: This article is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the author and not those of the U.S. Census Bureau. All time series analyzed in this article are from public data sources.

3. applicability to diverse sampling frequencies,
4. applicability to multiple frequencies of latent seasonality, which is potentially of non-integer period, and
5. ability to address over- and under-adjustment.

Currently available diagnostics satisfy some of these criteria, but none address all facets.

With regard to the first point, a rigorous quantification of Type I and II errors is needed so that in a production environment with thousands of time series, an analyst can set tolerance levels for quality control. The distribution theory for diagnostic statistics should be developed under a broad set of time series data generating processes, so that critical values are approximately germane for typical sample sizes. As to the second point, we intend that a diagnostic takes a low value if and only if seasonality is present in a stochastic process; if low values could be obtained by non- seasonal processes, or high values could be generated by seasonal processes, then the diagnostic will be worse than useless – because it will generate false (spurious) conclusions. Thirdly, we observe that diagnostics developed for X-12-ARIMA (Findley et al. 1998) were intended for monthly data, with some extensions possible for quarterly series. However, there is an increasing public demand for the publication of weekly and daily time series (see McElroy et al. 2018 for an overview of the applications of daily time series to understanding retail holiday patterns), which presents new challenges for official statistics; moreover, multiple forms of seasonality (point number four) can be present. For example, daily time series have a weekly effect and an annual effect, which are really just two different types of seasonal effects. Note that for daily data, the annual period is 365.25, and monthly effects have an average period of 30.25 – these non-integer seasonal periods create new challenges for diagnostics based upon seasonal autocorrelations.

The fifth point is concerned with over- and under-adjustment of time series, for which there is a long literature: Granger (1978) noted with concern the introduction of negative seasonal autocorrelation into a time series by application of the Wiener-Kolmogorov (Bell 1984) seasonal adjustment filters, although the phenomenon had been already described in Nerlove (1964). Also see Sims (1978), Tukey (1978), and Bell and Hillmer (1984). If the extraction of seasonality involves using overly stable seasonal filters, then seasonality will remain – we refer to this as *under-adjustment*. On the other hand, using overly dynamic seasonal filters produces negative seasonal correlation – we refer to this as *over-adjustment*. Ansley and Wecker (1984) and McElroy (2012) discuss a method that reduces over-adjustment, while model-based diagnostics of under- and over-adjustment are described in Maravall (2003), McElroy (2008), and Blakely and McElroy (2017). Whereas under-adjustment is clearly a problem – since measurable seasonality remains (possibly marring interpretations of growth rates) – over-adjustment may also be undesirable, because it indicates that non-seasonal dynamics (such as the business cycle) may have been removed from the data and erroneously allocated to the seasonal component. This is akin to the problem of trend extraction: under-smoothing means the extracted trend will have too many oscillations, whereas over-smoothing will force long-term trend movements into the business cycle.

This article focuses on proposing a test for over- or under-adjustment, while allowing for non-integer periods of seasonality. First, we must clearly parse the phenomenon of seasonality. We propose – based on ideas developed in Lin et al. (2019) – the following definition of seasonality: persistency in a time series over seasonal periods that is not explainable by intervening time periods. For a monthly series with a seasonal period equal to twelve, seasonality is indicated by persistency from year to year that is not explained by month-to-month changes. Note that both parts of this definition are crucial: without seasonal persistency from year to year, no seasonal pattern will be apparent, so this facet is clearly necessary; however, any trending time series also has persistency from year to year, which comes through the intervening months – we need to screen out such cases.

If the seasonality is non-stationary, there are diagnostic tools available in the econometric literature. Seasonal unit root tests (Hylleberg 1986; Hylleberg et al. 1990; Canova and Hansen 1995; Busetti and Harvey 2003) adopt either as null or alternative hypothesis that the form of persistence resembles a random walk year to year, for each season. Another framework involves periodically integrated processes (Franses 1994). Although such tests satisfy our first three criteria, these methods are not easily adapted to non-integer periods, and cannot address the over-adjustment problem; neither are they effective for diagnosing milder, dynamic seasonality that can be present in stationary processes. This latter application is vital for the detection of residual seasonality in seasonally adjusted data – seasonality in such series will not manifest unit roots, but rather a highly evolutive pattern that is consistent with a stationary formulation.

If a time series is covariance stationary, it is natural to parse persistency in terms of autocorrelation (cf. Proietti (1996), which measures the strength of autocorrelation at seasonal lags). As we show in this article, we can adapt persistency to non-integer lags of the autocovariance function (acvf) via its decomposition in terms of autoregressive (AR) roots, and examine seasonality of arbitrary frequency through the modulus and phase of the root. Whereas under-adjustment would be indicated by the presence of AR roots of near-unit magnitude and seasonal phase, over-adjustment corresponds to a negative form of persistency (i.e., negative seasonal autocorrelations) termed anti-persistency, and can be measured through moving average (MA) roots computed from the inverse autocovariances (McElroy and Roy 2018), that is, the autocovariances of the spectral density's reciprocal.

This framework of using AR roots as a diagnostic of under-adjustment (and MA roots for over-adjustment) satisfies the five criteria listed above. In Section 2 we develop the asymptotic theory and hypothesis testing framework, also demonstrating that small values of the diagnostic occur if and only if seasonality is present. (Appendix A of the Online Supplement offers a foundation for understanding oscillations and seasonality.) Section 3 applies the methodology, providing the implementation details; because seasonality becomes associated with the phase of AR roots, we can address arbitrary (regular) sampling frequencies and multiple non-integer period seasonalities. Simulation studies are given in Section 4, and data illustrations in Section 5. Section 6 concludes, with proofs and an additional illustration in the Supplement.

2. Methodology

2.1. Framework

Consider a weakly stationary process $\{X_t\}$ that is mean zero and purely unpredictable; then by the Wold Decomposition (Theorem 7.6.4 of McElroy and Politis 2020) $X_t =$

$\sum_{j=0}^{\infty}\psi_j Z_{t-j}$ for a white noise sequence $\{Z_t\}$, and the $\{\psi_j\}$ are called the Wold coefficients. Oscillatory behavior in the Wold coefficients corresponds to seasonality, as discussed in online Appendix A. That background discussion also shows that the oscillations of a sequence are governed by the magnitude and phase of the roots of the z-transform, that is, $\psi(z) = \sum_{j=0}^{\infty}\psi_j z^j$. In particular, large values of $|\psi(\rho^{-1}e^{i\omega})|$ for $\rho \in (0, 1)$ correspond to an oscillatory pattern in the ψ_j, where the frequency of the oscillation is ω and the pattern is "damped" (i.e., it decays as $j \to \infty$) by the value ρ. The higher the value of ρ, the slower the decay in the coefficients, resulting in a more persistent oscillation.

Consider a seasonal pattern of period s (i.e., the number of seasons per year), which has frequency $\omega = 2\pi/s$. Writing $\pi(z) = 1/\psi(z)$ (which converges outside the unit circle if the process is invertible), we say there is ρ-persistent seasonality of frequency ω if $\pi(\rho^{-1}e^{i\omega}) = 0$. In the special case that $\{X_t\}$ is an ARMA process with MA polynomial $\theta(z)$ and AR polynomial $\phi(z)$, such a seasonal pattern exists if $\phi(\rho^{-1}e^{i\omega}) = 0$, because $\pi(z) = \phi(z)/\theta(z)$ (see Theorem 5.5.3 of McElroy and Politis 2020). Similarly, oscillatory patterns in the coefficients $\{\pi_j\}$ corresponding to $\pi(z) = \sum_{j=0}^{\infty}\psi_j z^j$ correspond to anti-seasonality (see discussion in onine Appendix A). We say there is ρ-persistent anti-seasonality of frequency ω if $\psi(\rho^{-1}e^{i\omega}) = 0$; in the case of an ARMA process, this occurs if $\theta(\rho^{-1}e^{i\omega}) = 0$, because $\psi(z) = \theta(z)/\phi(z)$ (see Theorem 5.4.3 of McElroy and Politis 2020).

For the remainder of the article we suppose that an invertible ARIMA model has been identified and fitted to a sample $X_1, \ldots, X_T$ of size T from the data process $\{X_t\}$. If the process is stationary, then no differencing is needed and we can fit an ARMA model; the AR(∞) representation of the data process is $\pi(z) = \phi(z)/\theta(z)$, where ϕ and θ are stable AR and MA polynomials (i.e., all of their roots are outside the unit circle) that are relatively prime. If the process requires differencing, then we allow the AR polynomial $\phi(z)$ to have roots on the unit circle. Let $\delta(z) = 1 - \sum_{j=1}^{d}\delta_j z^j$ be the unit root portion of the polynomial $\phi(z)$, and let $\varphi(z)$ (of degree p) correspond to the non-unit roots. Then the pseudo-autoregressive polynomial $\phi(z)$ is defined as $\delta(z)\ \varphi(z)$. Hence, if we difference the data with $\delta(B)$ (where B is the backward shift operator), the resulting process is a stationary ARMA with AR polynomial $\varphi(z)$ and MA polynomial $\theta(z)$.

2.2. *Testing for Seasonality*

Whether or not there are unit roots in $\phi(z)$, seasonality can be tested in terms of the polynomial $\phi(z)$, because $\pi(z) = \phi(z)/\theta(z)$. In particular, for any given ω, the null hypothesis – that ρ_0-persistent seasonality of frequency ω is present – can be formulated as

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

Note that $H_0(\rho_0)$ holds if and only if $\phi(r^{-1}e^{i\omega}) = 0$ for some $r = \rho_0$. We can measure departures from ρ_0-persistent seasonality by computing $|\pi(\rho_0^{-1}e^{i\omega})|^2$, or its estimate based upon maximum likelihood estimation (MLE) of the ARMA parameters. Alternatively, in the case of an AR model we can compute ordinary least square estimates (OLS) of the parameters; the asymptotic theory is the same. Let $g(r) = |\pi(r^{-1}e^{i\omega})|^2$, and set

$$\hat{g}(r) = |\hat{\pi}(r^{-1}e^{i\omega})|^2, \quad (2)$$

where $\hat{\pi}(z) = \hat{\phi}(z)/\hat{\theta}(z)$ and the polynomials are estimated by replacing the coefficients with MLEs. Our test statistic of $H_0(\rho_0)$ is $T\hat{g}(\rho_0)$, where the rate T is justified by the subsequent asymptotic theory. Next, we present theory for this test statistic in the case that there are no unit roots in $\phi(z)$.

THEOREM 1. *Let $\{X_t\}$ be a causal invertible ARMA(p,q) process with AR polynomial $\phi(z)$ and MA polynomial $\theta(z)$. Let $\pi(z) = \phi(z)/\theta(z)$ and $\hat{g}(r)$ defined via (2), where the estimates for the ARMA parameters are obtained from a sample of size T via either MLE or OLS (in the pure AR case). It follows that when $g(r) = 0$*

$$T\hat{g}(r) \overset{\mathcal{L}}{\Rightarrow} \frac{|\underline{Z}'\underline{\zeta}|^2}{|\theta(r^{-1}e^{i\omega})|^2},$$

where $\underline{\zeta}_j = (re^{i\omega})^{-j}$ for $1 \leq j \leq p$ and $\underline{Z} \sim \mathcal{N}(0, \Gamma_p^{-1})$ such that Γ_p is the $p \times p$ Toeplitz covariance matrix corresponding to spectral density $|\phi(e^{-i\lambda})|^{-2}$. When $g(r) > 0$, instead

$$\sqrt{T}(\hat{g}(r) - g(r)) \overset{\mathcal{L}}{\Rightarrow} \mathcal{N}(0, V),$$

where $V = \underline{\eta}' F^{-1} \underline{\eta}$, F is the Fisher information matrix for the ARMA process (described in the proof), and

$$\underline{\eta} = \begin{bmatrix} -(\phi(r^{-1}e^{i\omega})\underline{\zeta} + \phi(r^{-1}e^{-i\omega})\underline{\bar{\zeta}})|\theta(r^{-1}e^{i\omega})|^{-2} \\ (\theta(r^{-1}e^{i\omega})\underline{\xi} + \theta(r^{-1}e^{-i\omega})\underline{\bar{\xi}})|\theta(r^{-1}e^{i\omega})|^{-4}|\phi(r^{-1}e^{i\omega})|^2 \end{bmatrix},$$

where $\underline{\xi}_j = (re^{i\omega})^{-j}$ for $1 \leq j \leq q$.

REMARK 1. The alternative hypothesis indicates that $g(\rho_0) > 0$, and Theorem 1 indicates that the test statistic is $O_P(T^{1/2})$ plus $T\,g(\rho_0)$ in that case, yielding a consistent test.

In cases where the data may have unit roots, a different theory is needed. If a pure AR model is fitted, one can use OLS, as this allows for the parameters corresponding to unit or explosive roots. (The Yule-Walker method (McElroy and Politis 2020), which enforces stability, should be avoided because of substantial bias when the process has roots close to the unit circle.) Alternatively, one can apply $\delta(B)$ to difference the data, and then fit a stationary AR model. If fitting an ARIMA model, this latter strategy is used: apply $\delta(B)$, and then fit via MLE an ARMA model. The following result allows us to test a null hypothesis of stationary seasonality ($\rho_0 < 1$) when unit roots are known to be present in the process – testing $\rho_0 = 1$ requires a more complicated limit theory that is only mentioned in the proof.

THEOREM 2. *Let $\{X_t\}$ be an invertible ARIMA($p + d$, q) process with differencing polynomial $\delta(B)$, stable AR polynomial $\varphi(z)$, and MA polynomial $\theta(z)$. Set $\phi(z) = \delta(z)\varphi(z)$, the pseudo- autoregressive polynomial. Let $\pi(z) = \phi(z)/\theta(z)$ and $\hat{g}(r)$ defined via (2), where the estimates for the ARMA parameters are obtained from a sample of size T via* *either MLE or OLS (in the pure AR case). It follows that when* $g(r) = 0$ *and* $r < 1$ *that*

$$T\hat{g}(r) \overset{\mathcal{L}}{\Rightarrow} \frac{|\underline{Z}'\underline{\zeta}|^2}{|\theta(r^{-1}e^{i\omega})|^2},$$

where $\underline{\zeta}_j = (re^{i\omega})^{-j}$ *for* $1 \leq j \leq p + d$ *and* $\underline{Z} \sim \mathcal{N}(0, P'\Gamma_p^{-1}P)$ *such that* Γ_p *is the* $p \times p$ *Toeplitz covariance matrix corresponding to spectral density* $|\varphi(e^{-i\lambda})|^{-2}$ *and* P *is a* $p \times p + d$*-dimensional matrix given by*

$$P = \begin{bmatrix} 1 & -\delta_1 & \ldots & -\delta_d & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 1 & -\delta_1 & \ldots & -\delta_d \end{bmatrix}. \tag{3}$$

When g(r) > 0, *instead*

$$\sqrt{T}(\hat{g}(r) - g(r)) \overset{\mathcal{L}}{\Rightarrow} \mathcal{N}(0, V),$$

where $V = \underline{\eta}'R'F^{-1}R\underline{\eta}$, $R =$ *diag{P,*I_q*} (here* I_q *is a q-dimensional identity matrix), F is the Fisher information matrix for the stationary ARMA process (described in the proof), and*

$$\underline{\eta} = \begin{bmatrix} -(\phi(r^{-1}e^{i\omega})\underline{\zeta} + \phi(r^{-1}e^{-i\omega})\bar{\underline{\zeta}})|\theta(r^{-1}e^{i\omega})|^{-2} \\ (\theta(r^{-1}e^{i\omega})\underline{\xi} + \theta(r^{-1}e^{-i\omega})\bar{\underline{\xi}})|\theta(r^{-1}e^{i\omega})|^{-4}|\phi(r^{-1}e^{i\omega})|^2 \end{bmatrix},$$

where $\underline{\xi}_j = (re^{i\omega})^{-j}$ *for* $1 \leq j \leq q$.

2.3. Testing for Anti-Seasonality

Suppose now that we wish to test for the presence of anti-seasonality. Now we wish to examine $\psi(\rho^{-1}e^{i\omega})$, and it is important that the process be stationary. Therefore suppose that $\{X_t\}$ is an ARMA process, where any non-stationary effects have been previously removed by a differencing polynomial $\delta(B)$. Then, for any given ω, the null hypothesis of ρ_0-persistent anti-seasonality is written

$$H_0(\rho_0) : \pi(r^{-1}e^{i\omega}) = 0 \,\text{has solution}\, r = \rho_0, \tag{4}$$

where $\psi(z) = \theta(z)/\phi(z)$. Here $H_0(\rho_0)$ holds if and only if $\theta(r^{-1}e^{i\omega}) = 0$ for some $r = \rho_0$. We can measure departures from ρ_0-persistent anti-seasonality by computing $\left|\psi\left(\rho_0^{-1}e^{i\omega}\right)\right|^2$; set $h(r) = |\psi(r^{-1}e^{i\omega})|^2$, and let

$$\hat{h}(r) = |\hat{\psi}(r^{-1}e^{i\omega})|^2. \tag{5}$$

Here $\hat{\psi}(z) = \hat{\theta}(z)/\hat{\phi}(z)$, and the polynomials are estimated by replacing the coefficients with MLEs. Theorem 1 can likewise be adapted by swapping the polynomials appropriately, as stated below. (The proof follows the same techniques, and is therefore omitted.) We emphasize that this theory requires an invertible moving average polynomial, and thus cannot be used to test an over-adjustment hypothesis where $\rho_0 = 1$.

COROLLARY 1. *Let $\{X_t\}$ be an invertible ARMA(p,q) process with AR polynomial $\phi(z)$ and MA polynomial $\theta(z)$. Let $\psi(z) = \theta(z)/\phi(z)$ and $\hat{h}(r)$ defined via (5), where the estimates for the ARMA parameters are obtained from a sample of size T via either MLE or OLS (in the pure AR case). It follows that when $h(r) = 0$*

$$T\hat{h}(r) \stackrel{\mathcal{L}}{\Rightarrow} \frac{|\underline{Z}'\underline{\xi}|^2}{|\phi(r^{-1}e^{i\omega})|^2},$$

where $\underline{\xi}_j = (re^{i\omega})^{-j}$ for $1 \leq j \leq q$ and $\underline{Z} \sim \mathcal{N}(0, \Gamma_q^{-1})$ such that Γ_q is the $q \times q$ Toeplitz covariance matrix corresponding to spectral density $|\theta(e^{-i\lambda})|^{-2}$. When $h(r) > 0$, instead

$$\sqrt{T}(\hat{h}(r) - h(r)) \stackrel{\mathcal{L}}{\Rightarrow} \mathcal{N}(0, V),$$

where $V = \underline{\eta}'F^{-1}\underline{\eta}$, F is the Fisher information matrix for the ARMA process, and

$$\underline{\eta} = \begin{bmatrix} -(\phi(r^{-1}e^{i\omega})\underline{\zeta} + \phi(r^{-1}e^{-i\omega})\bar{\underline{\zeta}})|\phi(r^{-1}e^{i\omega})|^{-4}|\theta(r^{-1}e^{i\omega})|^2 \\ (\theta(r^{-1}e^{i\omega})\underline{\xi} + \theta(r^{-1}e^{-i\omega})\bar{\underline{\xi}})|\phi(r^{-1}e^{i\omega})|^{-2} \end{bmatrix},$$

where$\underline{\zeta}_j = (re^{i\omega})^{-j}$ for $1 \leq j \leq p$.

A limitation of the methodology behind Corollary 1 is that it cannot be applied to non-invertible processes; to rectify this, we proceed by considering instead of $h(r)$ the related quantity

$$f(z) = \sum_{|h| \leq q} \gamma_h z^h = \psi(z)\psi\left(z^{-1}\right)\sigma^2 \tag{6}$$

for $z \in \mathbb{C}$, where $\gamma_h = \mathrm{Cov}\left[X_{t+h}, X_t\right]$ is the autocovariance function of the stationary process $\{X_t\}$. We evaluate Equation (6) at $z = r^{-1}e^{i\omega}$. Although $f(r^{-1}e^{i\omega})$ is not equal to $h(r)$ (in fact, it is complex-valued when $r \neq 1$), we still have $f(r^{-1}e^{i\omega}) = 0$ if and only if $\psi(r^{-1}e^{i\omega}) = 0$. This suggests basing a test statistic on

$$\hat{f}(z) = \sum_{|h| \leq q} \hat{\gamma}_h z^h, \tag{7}$$

where $\hat{\gamma}_h$ is the sample autocovariance based on a sample of size T. Although $\hat{f}(z)$ is not real-valued (and need not be positive-definite when $z = e^{-i\lambda}$, because the choice of q implicitly generates a truncation taper), a distribution theory is easily developed based upon the sample autocovariances, and a central limit theorem can be established. Hence, we propose to test the null $H_0(\rho_0)$ of anti-seasonality with $T\left|\hat{f}\left(\rho_0^{-1}e^{i\omega}\right)\right|^2$. The following theory describes the asymptotic distribution, and allows for $r = 1$ without any qualitative change to the results.

THEOREM 3. *Let $\{X_t\}$ be a possibly non-invertible MA(q) process with independent and identically distributed inputs and moving average polynomial $\theta(z)$. With $f(z)$ and $\hat{f}(z)$ defined via Equations (6) and (7) for any $z \in \mathbb{C}$,*

$$T\left|\hat{f}(z) - f(z)\right|^2 \stackrel{\mathcal{L}}{\Rightarrow} |\underline{Z}'\underline{\upsilon}|^2,$$

where $\underline{v}' = \left[1, z + z^{-1}, \ldots, z^q + z^{-q}\right]$ *and* $\underline{Z} \sim \mathcal{N}\left(0, V\right)$ *such that V is the* $q + 1 \times q + 1$ *asymptotic covariance matrix of the sample autocovariances at lags 0 through q, that is, the jkth entry (for* $0 \leq j, k \leq q$*) is given by*

$$V_{jk} = \frac{2}{2\pi} \int_{-\pi}^{\pi} \cos(\lambda j) \cos(\lambda k) f(e^{-i\lambda})^2 d\lambda$$

$$+ \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \cos(\lambda j) \cos(\omega k) \Omega(e^{i\lambda}, e^{-i\omega}, e^{i\omega}) d\lambda d\omega,$$

where Ω *is the tri-spectral density given in (B.2).*

Remark 2. Under the null hypothesis, with $z = \rho_0^{-1} e^{i\omega}$, the test statistic converges to $|\underline{Z}'\underline{v}|^2$, whose critical values can be simulated. If the inputs have zero kurtosis then the tri-spectral density is zero, and we can estimate V by utilizing $\hat{f}\left(e^{-i\lambda}\right)$ as a plug-in estimator of $f(e^{-i\lambda})$, the spectral density; alternatively, we can plug in the periodogram and divide by 2, approximating the integral by a Riemann sum over Fourier frequencies – see Chiu (1988) and Deo and Chen (2000). We find that in simulations, the size of this latter approach is far superior. The alternative hypothesis indicates that $f\left(\rho_0^{-1} e^{i\omega}\right)$ is non-zero; asymptotically the test statistic equals $T\left|f\left(\rho_0^{-1} e^{i\omega}\right)\right|^2$ plus lower-order terms, and this quantity will generate power (because it is $O(T)$).

3. Applications and Implementation

In the practice of seasonal adjustment one of the key tasks is to decide whether a given time series should be seasonally adjusted. Secondarily, series that have been seasonally adjusted should be assessed for adequacy, and among the potential problems it is important to determine whether the series have been under-adjusted (or over-adjusted). As discussed in the Introduction (with further exposition in online Appendix A), under-adjustment is characterized by the presence of dynamic seasonality, whereas over-adjustment is characterized by the presence of anti-seasonality. Therefore, there are three potential applications of the testing methodology of Section 2:

1. a test for residual seasonality (i.e., a test of under-adjustment), used upon stationary data that has already been adjusted (or clearly has no unit roots) and trend-differenced, if needed,
2. a test for raw seasonality, used upon potentially non-stationary data with unit roots, and
3. a test of over-adjustment, used upon seasonally adjusted data where there is concern about residual anti-seasonality.

For these three cases, we propose using the test statistics discussed in Theorem 1, Theorem 2, and Theorem 3 respectively, using the distribution theory for the null hypothesis to obtain critical values. In particular, cases one and two utilize the test statistic $T\hat{g}\left(\rho_0\right)$ given by (2) to test $H_0(\rho_0)$ given by (1). The third case utilizes the test statistic $T\left|\hat{f}\left(\rho_0^{-1} e^{i\omega}\right)\right|^2$ given by (7) to test $H_0(\rho_0)$ given by (4). Note that these null hypotheses can be calibrated according to the concerns and priorities of the seasonal adjuster, through the determination of ρ_0.

All of these tests are upper one-sided, with large values of the test statistic indicating rejection of $H_0(\rho_0)$. However, we do not have a way of knowing whether the true seasonal persistence (or anti-persistence) – if it exists, because possibly $\pi\left(r^{-1}e^{i\omega}\right) \neq 0$ for all r – is greater than or less than ρ_0. To address this question, we can consider computing both the test statistic and p-values for a range of values of $\rho_0 \in (0,1)$. Because both the test statistic and the critical values are continuous functions of ρ_0, we thereby obtain p-values as a continuous function of ρ_0, say $p(\rho_0)$. If desired, we can invert this function, for example obtaining $p^{-1}(\alpha,1)$ as the open set of ρ_0 such that we fail to reject $H_0(\rho_0)$ at confidence level α. Of course, it may happen that the p-values are low for all values of ρ_0, and such intervals are empty; then there is no seasonality of any degree of persistence present. For other applications, it is useful to plot $p(\rho_0)$ for $\rho_0 \in (0,1)$, or for ρ in some sub-interval of (0,1) that corresponds to the degrees of seasonal persistency that are deemed to be of interest. If $p(\rho) < \alpha$ for all ρ_0 in the given interval, then we can reject seasonality at all those persistencies.

If it is desired to obtain a joint test over J different frequencies $\omega_1, \ldots, \omega_J$, then the following approach can be used. We may have a null hypothesis for each frequency ω_j of persistency $\rho_0^{(j)}$, and we say that the null hypothesis for seasonality across all frequencies holds if and only if all the individual null hypotheses are valid:

$$H_0\left(\rho_0^{(1)}, \ldots, \rho_0^{(J)}\right) : \pi\left(\rho_0^{(j)} e^{i\omega_j}\right) = 0 \quad \text{for all } 1 \leq j \leq J. \tag{8}$$

(In the case of testing for anti-seasonality, we examine $\psi\left(\rho_0^{(j)} e^{i\omega_j}\right)$ instead of $\pi\left(\rho_0^{(j)} e^{i\omega_j}\right)$.) By taking the minimum of the various test statistics, a rejection (of seasonality) occurs if and only if all of the individual test statistics are significant; we fail to reject if at least one test statistic is small, that is, there is at least one j for which $\rho_0^{(j)}$-persistent seasonality exists at frequency ω_j. This indicates that the minimum is an appropriate statistic; for testing seasonality we have the joint test statistic

$$\min_{1 \leq j \leq J} T|\hat{\pi}(e^{i\omega_j}/\rho_0^{(j)})|^2, \tag{9}$$

whereas for anti-seasonality testing we use

$$\min_{1 \leq j \leq J} T|\hat{\psi}(e^{i\omega_j}/\rho_0^{(j)})|^2. \tag{10}$$

The critical values are easily obtained by simulation, as the results of Theorems 1, 2, and 3 are clearly joint across various ω_j. If we set each $\rho_0^{(j)}$ equal to a common ρ_0 for each of the component null hypotheses, then we obtain a test statistic and critical value as a function of a single number ρ_0, and hence we can obtain $p(\rho_0)$ in the manner described for a single frequency.

Each application will dictate the frequencies ω that should be considered. For a time series with s seasons per year (or other appropriate unit of time), one should examine $\omega = 2\pi j/s$ for $1 \leq j \leq \lfloor s/2 \rfloor$. For monthly data there are six frequencies to test; for quarterly data there are two frequencies to test. (Though commonly, the last frequencies – which are equal to π – are ignored, so that either five or one frequencies are actually considered for monthly and quarterly data.) A daily series has both a weekly and annual periodicity, and hence $\omega = 2\pi j/7$ for $1 \leq j \leq 3$ and $\omega = 2\pi j/365.25$ for $1 \leq j \leq 182$ are potential frequencies of interest. Given the large number of test statistics with a joint dependence structure – which, however, test different null hypotheses – some discretion is needed when combining into a single test of seasonality, as the significance levels may become modified.

In practice, when testing for seasonality we must specify an ARMA model and fit it via MLE. Our own implementation instead fits a high order AR(p) model via OLS, with p selected via some empirical means – such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) discussed in McElroy and Politis (2020). When the order p is over-specified, but fixed, the test statistics are still asymptotically correctly sized – though there may be a loss of power in finite samples, due to the inefficiency of estimating some AR coefficients that are truly zero. However, if p is allowed to grow with sample size (as with the use of AIC and BIC), then there is some probability of getting a high order model, which unduly adds to the statistic's variability, thereby lowering power. On the other hand, a value of p that is too small implies a mis-specification, causing bias in the test statistics. Whereas BIC is consistent for model order, AIC is upward-biased, tending to over-specify models. Because in this particular case over-specification is less hazardous to the asymptotic theory, the case can be made that AIC is preferable – and this surmise is borne out by our simulation studies.

When testing for anti-seasonality we need to determine the order q of the MA representation of the differenced seasonally adjusted data. Again, an information criterion could be used in conjunction with maxmimum likelihood fitting of the MA model – however, in the case that $\rho_0 = 1$ the moving average process is non-invertible and the usual asymptotic theory upon which AIC and BIC are founded cannot be applied (Tanaka 1996). Alternatively, we can let $\hat{q}$ be the largest q such that the sample autocovariances at all lags higher than q fail to reject the null hypothesis that the corresponding autocovariance is zero; we can use the variance estimate given in Paradigm 10.1.2 of McElroy and Politis (2020) to obtain a studentized test statistic. In our implementation we use normal critical values corresponding to a significance level of $1/\sqrt{T/3}$, which allow for the test's size to dwindle as sample size increases – thereby ensuring that q increases with T in an empirical fashion.

A final consideration is that components – such as the seasonal adjustment or the irregular – that are the output of a seasonal adjustment procedure will typically have nonlinear distortions in the beginning and final portions of the sample, due to forecast extension and/or adaptive filters; because the impact of forecast extension is localized to the edges of the sample, one can trim the component of the first and last 2 to 3 years – see Findley et al. (1998). In summary, we propose the following sequential procedure for de-seasonalizing a time series:

1. Test the raw series for seasonality by fitting an ARIMA model without seasonal differencing, testing the null (8) for $\rho_0 \in (.97, 1)$ with test statistic (9) for desired seasonal frequencies,
2. If dynamic seasonality is present (indicated by failure to reject in step 1 above, for at least one $\rho_0 \in (.97, 1)$), seasonally adjust the time series and proceed to steps 3 and 4 (otherwise, the procedure is complete),
3. Test the seasonally adjusted series for seasonality by trimming the first and last three years of data (to reduce the impact of non-linearity at the sample boundaries) and

fitting an ARIMA model, testing the null (8) for $\rho_0 \in (.9, 1)$ with test statistic (9) for desired seasonal frequencies, and
4. Test the seasonally adjusted series for anti-seasonality by trimming the first and last three years of data (to reduce the impact of non-linearity at the sample boundaries) and testing the joint null – given by swapping ψ for π in (8) – for $\rho_0 \in (.5, 1)$ with test statistic (10) for desired seasonal frequencies.

We remark that critical values for tests are generated through Monte Carlo simulation, using the asymptotic distributions given in Section 2; R code is available from the author. The recommended seasonal frequencies are $\pi/2$ for quarterly data, and $\omega_j = \pi j/6$ (for $1 \leq j \leq 5$) for monthly data. The range of ρ_0 considered in the seasonality tests for raw data are taken fairly high, the interval (.97, 1) being a suggestion based on the discussion in online Appendix A; the idea here is that seasonal adjustment should not be undertaken unless a fairly substantial degree of dynamic seasonality is present. However, some practitioners have suggested seasonally adjusting time series in which a much milder degree of seasonality is present (i.e., broaden the range of ρ_0), to mitigate the possibility of seasonality manifesting in aggregations of seasonally adjusted series – which has been documented in Moulton and Cowan (2016).

In step 3, we instead consider a broader interval of (.9, 1), so that we are more concerned about the presence of dynamic seasonality in the adjustment, as compared to the raw. Alternatively, one could make this range consistent with step 1, in this example setting the interval to (.97, 1). Similarly, in step 4 we screen out even mild cases of anti-seasonality by setting the interval to (.5, 1). Finally, we note that some practitioners may not be concerned about the presence of anti- seasonality, since it is a necessary outcome of model-based seasonal adjustment; such users could just omit step 4.

After completing these four steps, either the analyst is satisfied with the outcome (i.e., either the series needs no adjustment, or it does, and its seasonal adjustment is deemed to be adequate) or there is some deficiency, e.g., under- or over-adjustment. In this latter case, the analyst may wish to re-examine the modeling of the time series – frequently, a different specification of outlier effects, or a change to the ARIMA model, can result in an improved seasonal adjustment. The case of over-adjustment is harder to address, possibly requiring a different set of seasonal adjustment filters that produce narrower seasonal troughs in the spectral density; this is in contrast to the case of under-adjustment, where a more dynamic filter (with a wider seasonal spectral trough) is needed.

4. Numerical Experiments

To discern the efficacy of the method in finite samples, we consider simulating from a few different processes.

4.1. Atomic Seasonality With Transient Noise

We study monthly Gaussian time series $\{X_t\}$ generated from an AR(3) model with AR polynomial

$$\phi(z) = (1 - \tau z)\big(1 - 2\rho \cos\big(\pi/6\big)z + \rho^2 z^2\big). \tag{11}$$

This process corresponds to an atomic seasonality (i.e., there is a single seasonal frequency involved) at the first seasonal frequency, muddled by the presence of a transient effect. The acvf and spectrum are plotted in Figure 1, where we have set $\tau = .8$ and $\rho = .9$. The true AR roots have magnitudes of 1.25 (for the real root, corresponding to the transient effect) and 1.11 (for the complex roots corresponding to the atomic seasonality). From the plots, it is apparent that the strong seasonality is somewhat attenuated by the transient effect, so the impact of the atomic seasonality is weaker than it would be if $\tau = 0$. As a second example, we lower the seasonal persistency to $\rho = .8$, and dampen the transient component by setting $\tau = .3$, displayed in Figure 2. Here the weak seasonality is apparent, no longer being obfuscated by a transient effect.

For both of these processes (Figures 1 and 2) we generate 10,000 Gaussian simulations for each sample size $T = 12n$, where n is the number of years and $n = 5, 10, 15, 20$. We apply our procedure with $\omega = \pi/6$ under four different scenarios: first, assuming that the AR order $p = 3$ is known in the calculation of the test statistic (based on fitting an AR(3) to each simulation) – in this case, we use critical values from the true AR(3) process. Second, a more realistic scenario determines the critical values from the AR(3) fitted to each simulation, but still assumes the true order $p = 3$ is known. Third, we use an over-specified order ($p = 24$, which is twice the number of seasons) to compute the test statistic and critical values. Fourth, we use AIC to identify the order p. We also explored the use of BIC and an AR identification rule given in McElroy and Politis (2020, 335), but these yielded much more badly mis-sized results and were not pursued further. In each case the critical

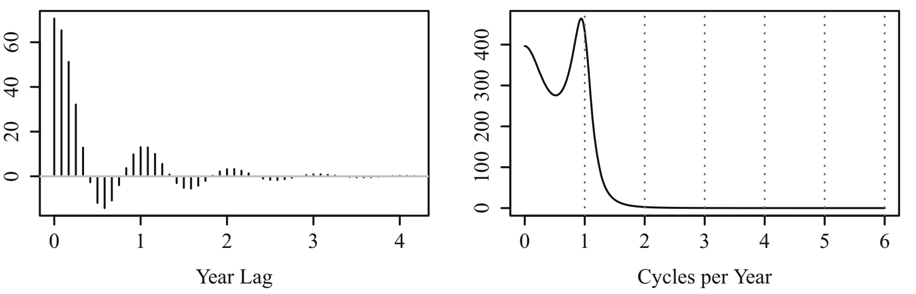

Fig. 1. Autocovariance function (left panel) and spectral density (right panel) for a seasonal AR(3) process ($\rho = .9$) with transient effect ($\tau = .8$). Autocovariance function is plotted as a function of lag divided by 12; spectral density is plotted as a function of cycles per year, or 12 divided by period.

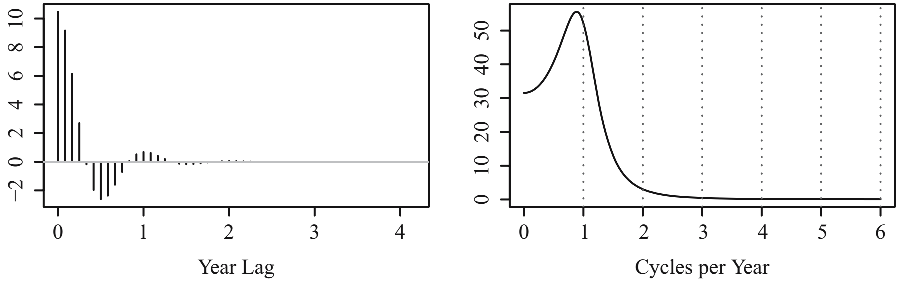

Fig. 2. Autocovariance function (left panel) and spectral density (right panel) for a seasonal AR(3) process ($\rho = .8$) with transient effect ($\tau = .3$). Autocovariance function is plotted as a function of lag divided by 12; spectral density is plotted as a function of cycles per year, or 12 divided by period.

values are generated for $\alpha = .10$, .05, .01; for the second two cases, we report the proportion of p-values that are less than α.

We examine the size of the procedure by taking as null hypothesis that $\rho = .9$ and $\rho = .8$ respectively, for the two processes. We also examine the power in both directions: with the first process, we adopt the null hypothesis that $\rho = .8$, and with the second process we adopt the null hypothesis that $\rho = .9$. The size and power results for the first process are given in Tables 1 and 3, where the alternative entails a greater degree ($\rho = .9$) of

Table 1. Size simulations from an AR(3) DGP (corresponding to Figure 1) based on a null hypothesis of .9-persistent seasonality at frequency $\pi/6$. Results are for known AR order (first three rows), for unknown parameters (second three rows), over-specified AR order (third three rows), and AIC-determined AR order (last three rows).

α	5 years	10 years	15 years	20 years
.10	.149	.116	.113	.101
.05	.092	.067	.061	.054
.01	.039	.026	.021	.018
.10	.139	.115	.108	.100
.05	.084	.059	.058	.048
.01	.028	.014	.012	.011
.10	.381	.206	.152	.140
.05	.274	.133	.089	.079
.01	.135	.050	.026	.020
.10	.625	.340	.183	.139
.05	.582	.297	.131	.086
.01	.508	.258	.087	.035

Table 2. Size simulations from an AR(3) DGP (corresponding to Figure 2) based on a null hypothesis of .8-persistent seasonality at frequency $\pi/6$. Results are for known AR order (first three rows), for unknown parameters (second three rows), over-specified AR order (third three rows), and AIC-determined AR order (last three rows).

α	5 years	10 years	15 years	20 years
.10	.135	.115	.109	.104
.05	.077	.058	.057	.053
.01	.021	.012	.013	.011
.10	.136	.114	.107	.103
.05	.074	.060	.057	.055
.01	.020	.014	.011	.012
.10	.424	.192	.152	.132
.05	.323	.117	.086	.072
.01	.157	.040	.023	.018
.10	.413	.483	.458	.377
.05	.297	.375	.374	.308
.01	.135	.182	.217	.192

Table 3. Power simulations from an AR(3) DGP (corresponding to Figure 1) with null hypothesis of .8-persistent seasonality at frequency π/6. Results are for known AR order (first three rows), for unknown parameters (second three rows), over-specified AR order (third three rows), and AIC- determined AR order (last three rows).

α	5 years	10 years	15 years	20 years
.10	.502	.758	.908	.961
.05	.304	.604	.820	.918
.01	.060	.230	.509	.719
.10	.465	.707	.852	.928
.05	.320	.565	.751	.865
.01	.107	.281	.476	.659
.10	.437	.207	.151	.134
.05	.328	.130	.087	.074
.01	.170	.045	.024	.020
.10	.787	.685	.677	.740
.05	.721	.578	.554	.625
.01	.611	.395	.321	.365

persistence than is hypothesized ($\rho = .8$). As for the second process, the size and power results are given in Tables 2 and 4, where the alternative entails a lesser degree ($\rho = .8$) of persistence than is hypothesized ($\rho = .9$).

In general, the size is over-estimated but may be considered adequate for 10 years of data when p is known; when p is fixed and over-specified, at least 20 years of data is needed. Moreover, the use of AIC results in badly mis-sized test statistics in the case of the second process, but AIC works adequately with 20 years of data in the case of the first

Table 4. Power simulations from an AR(3) DGP (corresponding to Figure 2) with null hypothesis of .9-persistent seasonality at frequency π/6. Results are for known AR order (first three rows), for unknown parameters (second three rows), over-specified AR order (third three rows), and AIC- determined AR order (last three rows).

α	5 years	10 years	15 years	20 years
.10	.322	.506	.659	.786
.05	.230	.386	.541	.681
.01	.108	.205	.321	.460
.10	.281	.503	.683	.804
.05	.181	.365	.554	.694
.01	.054	.149	.292	.440
.10	.409	.204	.160	.142
.05	.306	.133	.092	.083
.01	.155	.049	.025	.023
.10	.248	.421	.605	.740
.05	.166	.318	.502	.644
.01	.062	.151	.285	.420

process. Power is good for both processes when p is known, but drops considerably when p is over-specified, as expected. This occurs because the additional variability due to over-specification overwhelms the $T\,g(\rho_0)$ quantity in small samples. In these two cases where H_a holds, $g(\rho_0)$ is given by .0048 and .0059 respectively.

4.2. *Multiple Seasonal Peaks*

Next, we consider a more nuanced process where there is apparent either a mild or intense degree of seasonality at one or more seasonal frequencies. The process $\{X_t\}$ is a monthly Gaussian AR(10) time series with AR polynomial defined by

$$\phi(z) = \prod_{j=1}^{5} \left(1 - 2\rho^{(j)} \cos(\pi j/6) z + \rho^{(j)^2} z^2\right). \tag{12}$$

As a result, $g(r) = 0$ for $r = \rho^{(j)}$, for each $\omega_j = \pi j/6$ $(1 \leq j \leq 5)$. In Table 5 we report size results for the process corresponding to $\rho_0^{(j)} = .9$ for $1 \leq j \leq 5$ (again with 10,000 simulations for each sample size), where the joint seasonality test is utilized by taking the minimum of the frequency-specific test statistics. As above, we consider three scenarios regarding the model order p, either using the true $p = 10$ or the over-specified $p = 24$. In each case the critical values are generated for $\alpha = .10, .05, .01$; for the second two cases, we report the proportion of p-values that are less than α.

For power, we first consider the same process but change the null hypothesis to $\rho_0^{(j)} = .97$, with results reported in Table 6. Here $g(.97)$ takes the values .399, .135, .102, .135, and .399 respectively for $\omega_j = \pi j/6$ $(1 \leq j \leq 5)$. This situation corresponds to where the actual degree of seasonality is milder than is hypothesized, with rejections indicating no seasonality is present. As a second study, we adopt this same null hypothesis but now the true process is non-stationary with mingled seasonal persistencies:

Table 5. Size simulations from a stationary AR(10) DGP with null hypothesis of .9-persistent seasonality at various frequencies $\pi j/6$, $1 \leq j \leq 5$. Results are for known AR order (first three rows), for unknown parameters (second three rows), over-specified AR order (third three rows), and AIC-determined AR order (last three rows).

α	5 years	10 years	15 years	20 years
.10	.124	.114	.105	.107
.05	.068	.054	.053	.053
.01	.017	.011	.012	.011
.10	.169	.129	.118	.107
.05	.104	.070	.063	.053
.01	.035	.017	.017	.012
.10	.395	.169	.139	.122
.05	.291	.099	.077	.064
.01	.147	.028	.021	.014
.10	.577	.271	.179	.177
.05	.477	.201	.117	.109
.01	.285	.112	.043	.035

Table 6. Power simulations from a stationary AR(10) DGP with null hypothesis of .97-persistent seasonality at various frequencies $\pi j/6$, $1 \leq j \leq 5$. Results are for known AR order (first three rows), for unknown parameters (second three rows), over-specified AR order (third three rows), and AIC-determined AR order (last three rows).

α	5 years	10 years	15 years	20 years
.10	.447	.818	.955	.990
.05	.333	.732	.922	.981
.01	.172	.537	.810	.936
.10	.509	.859	.967	.994
.05	.394	.778	.941	.986
.01	.195	.574	.842	.950
.10	.427	.394	.531	.660
.05	.330	.294	.421	.560
.01	.178	.143	.231	.353
.10	.776	.838	.942	.979
.05	.737	.762	.907	.967
.01	.635	.577	.794	.919

$\rho_0^{(1)} = \rho_0^{(2)} = \rho_0^{(3)} = 1$, $\rho_0^{(4)} = \rho_0^{(5)} = .9$. The results are reported in Table 7; we omit entries for the case of known parameters, since the true process is non-stationary. The values of the functional $g(.97)$ are .121, .040, .031, .247, and .742 respectively for $\omega_j = \pi j/6$ $(1 \leq j \leq 5)$. This situation corresponds to where the actual degree of seasonality is greater than is hypothesized for the first three seasonal frequencies, with rejections indicating strong seasonality is present.

The size results are similar to the cases of atomic seasonality considered previously: when the AR order is known, at least 10 years of data is needed, but 20 years may be needed if one uses AIC or an over-specified order. Power against the stationary alternative is above 50% with 10 years of data if the order is known, but there is a loss of power in the

Table 7. Power simulations from a non-stationary AR(10) DGP with null hypothesis of .97- persistent seasonality at various frequencies $\pi j/6$, $1 \leq j \leq 5$. Results are for unknown parameters (first three rows), over-specified AR order (second three rows), and AIC-determined AR order (last three rows).

α	5 years	10 years	15 years	20 years
.10	.939	.988	.996	.999
.05	.918	.982	.996	.998
.01	.866	.970	.991	.997
.10	.738	.863	.939	.965
.05	.660	.803	.913	.951
.01	.478	.661	.841	.915
.10	.912	.979	.991	.995
.05	.883	.971	.986	.993
.01	.805	.948	.975	.989

over-specified case (as expected). Therefore the results for multiple peaks are qualitatively similar to those of a single peak. For the case of a non-stationary alternative, the power is much higher, even in the over-specified case; this indicates it is easier to differentiate between non-stationary seasonality and strong stationary seasonality, versus discriminating between various cases of stationary seasonality.

4.3. *Completely Non-Seasonal*

Next, consider the case of an AR(1) process of parameter .8, which we suppose is observed as a quarterly series. Because this process is clearly non-seasonal, we should expect our AR diagnostic to have high power when we set the null hypothesis at a moderate degree of seasonality. In this case we set $\rho_0 = .9$ for the null, and consider frequency $\omega = \pi/2$, the quarterly frequency (again generating 10,000 simulations for each sample size). For comparison, we also investigate the Q_s statistic of Maravall (Findley et al. 2017): because the AR(1) process is non-seasonal, we expect the proportion of p-values less than a given α to be approximately α. However, the actual autocorrelations are $\rho_4 = .41$ and $\rho_8 = .17$, which are substantially different from zero – and Q_s is predicated upon $\rho_4 = \rho_8 \leq 0$ as an appropriate metric (i.e., necessary and sufficient) for non-seasonality. Table 8 indicates that Q_s is mis-sized with five years of data, and has size surpassing 50% as the sample size increases. This merely illustrates that Q_s tends to flag such non-seasonal processes as seasonal, because it fails to account for seasonal lag correlation that is explained through the linkages of the other seasons. By way of contrast, the AR diagnostic has 100% power for all the settings given in Table 8, demonstrating that the new test correctly classifies the AR(1) process as non-seasonal.

4.4. *Testing for Over-Adjustment*

To study the test for over-adjustment, we consider a process with anti-seasonality at frequency $\pi/6$, that is, such that $f(\rho^{-1} e^{i\pi/6}) = 0$. Such a condition is satisfied by an MA(2) model with MA polynomial

$$\theta(z) = 1 - 2\rho \cos\left(\pi/6\right) z + \rho^2 z^2.$$

The inverse acvf and spectrum are plotted in Figure 3, where we have set $\rho = .9$. The true MA roots have magnitudes of 1.11 (for the complex roots corresponding to the anti-seasonality). From the plots, it is apparent that the strong anti-seasonality exists due to the trough, as well as the oscillations of the inverse acvf.

We generate 10,000 Gaussian simulations of each sample size $T = 12n$, for $n = 5, 10, 15, 20$, and apply the over-adjustment test based upon the sample autocovariances. For

Table 8. Size simulations from an AR(1) DGP for the Q_s diagnostic, based on a null hypothesis of no seasonality.

α	5 years	10 years	15 years	20 years
.10	.012	.286	.509	.660
.05	.000	.225	.436	.590
.01	.000	.125	.300	.456

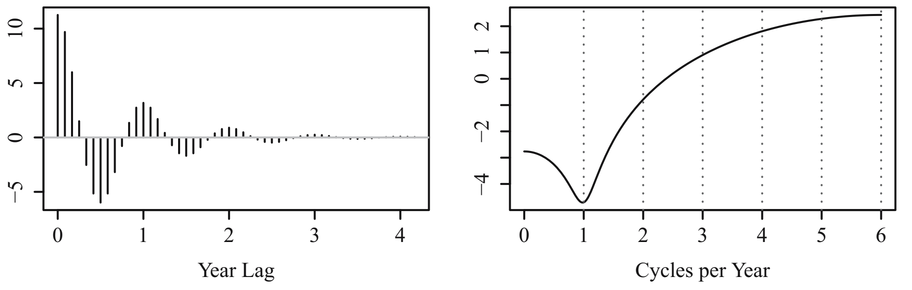

Fig. 3. Inverse autocovariance function (left panel) and log spectral density (right panel) for an anti-seasonal MA(2) process ($\rho = .9$). Inverse autocovariance function is plotted as a function of lag divided by 12; log spectral density is plotted as a function of cycles per year, or 12 divided by period.

scenarios, we suppose that the MA order $p = 2$ is known in the calculation of the test statistic, and we use critical values from the true MA(2) process. Second, we obtain critical values for the specified MA(2) with autocovariances fitted to each simulation. Third, we use an over-specified order ($q = 12$, which equals the number of seasons) to compute the test statistic and critical values. Finally, we determine the order q empirically by the procedure mentioned in Section 3. In each case the critical values are generated for $\alpha = .10,.05,.01$; for the second two cases, we report the proportion of p-values that are less than α.

We examine the size of the procedure by taking as null hypothesis that $\rho = .9$, and then investigate power by changing the process to having less anti-seasonality ($\rho = .5$). The size results are given in Table 9, and the power results are in Table 10. The test is slightly conservative, but for the first two cases (where the MA order is known) the size is adequate if there are 15 years of data. There is some deterioration in size with the third case (where q is over-specified), but results are not much worse when q is determined empirically. As for power, the gap between true and hypothesized ρ is fairly large, and the

Table 9. Size simulations from a MA(2) DGP based on a null hypothesis of .9-persistent anti- seasonality at frequency $\pi/6$. Results are for known MA order (first three rows), for unknown parameters (second three rows), for over-specified MA order (third three rows), and for empirically- determined MA order (final three rows).

α	5 years	10 years	15 years	20 years
.10	.084	.086	.096	.092
.05	.044	.046	.050	.047
.01	.013	.012	.012	.011
.10	.069	.080	.089	.094
.05	.022	.033	.036	.047
.01	.002	.003	.005	.007
.10	.047	.067	.074	.084
.05	.016	.025	.030	.037
.01	.001	.002	.003	.004
.10	.046	.056	.075	.091
.05	.018	.019	.023	.022
.01	.003	.003	.003	.002

Table 10. Power simulations from a MA(2) DGP with $\rho = .5$, based on a null hypothesis of .9- persistent anti-seasonality at frequency $\pi/6$. Results are for known MA order (first three rows), for unknown parameters (second three rows), for over-specified MA order (third three rows), and for empirically-determined MA order (final three rows).

α	5 years	10 years	15 years	20 years
.10	.373	.630	.829	.921
.05	.239	.465	.682	.821
.01	.085	.205	.369	.517
.10	.413	.671	.834	.921
.05	.267	.512	.697	.826
.01	.085	.233	.381	.563
.10	.057	.093	.117	.128
.05	.022	.041	.057	.069
.01	.000	.005	.009	.013
.10	.057	.066	.087	.111
.05	.021	.021	.022	.030
.01	.002	.001	.002	.002

results are adequate when model order is known. However, when q is over-specified or empirically determined, the variability in the estimation of V is excessive and the power is fairly low.

5. Empirical Illustrations

This section contains several empirical examples. First, we study a fairly typical monthly retail series and apply the testing procedure described in Section 3. Second, we test several raw daily time series for seasonality, illustrating the capability of the new diagnostics to handle different seasonal frequencies. Then, we consider the problem of residual seasonality in published US GDP. Finally, we compare two types of seasonal adjustments for a large collection of published monthly time series, and compare the diagnostic results to those obtained using Q_s. An additional illustration is included in Appendix C of the online Supplement.

5.1. Retail 442

We first analyze series 442 (Furniture and Home Furnishings Stores) of Retail Trade and Food Services, U.S. Census Bureau, covering the sample period of January 1992 through August 2019 (see Figure 4). We test the raw series for seasonality, and then seasonally adjust the series using the X11 method of X-13ARIMA-SEATS (U.S. Census Bureau 2015), which involves identification of outliers, holiday effects, and a SARIMA model for forecast extension of the data. Next, we test the seasonal adjustment for both under-adjustment and over-adjustment. Here we focus on producing a confidence interval for ρ, for each of the five seasonal frequencies $\pi j/6$ ($1 \leq j \leq 5$), for the cases considered.

The series is displayed with its seasonal adjustment (left panel) and seasonal factors (right panel) of Figure 4. The left panel seems to indicate a very obvious seasonal pattern, which is confirmed by the seasonal factors displayed in the right panel. To apply the test

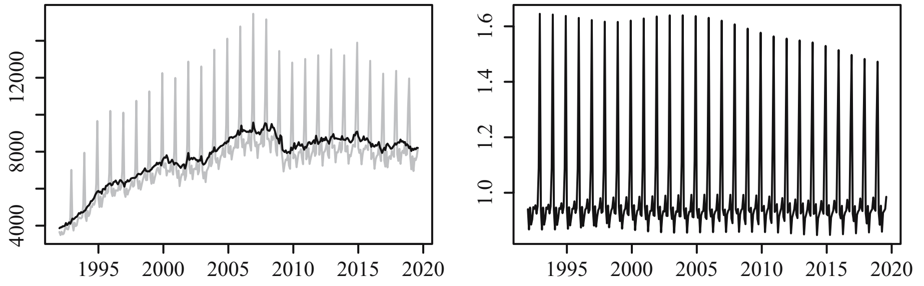

Fig. 4. Retail series 442 (furniture and home furnishings stores): raw series in grey with seasonal adjustment in black (left panel), and seasonal factors in black (right panel).

statistic in practice, we fit the AR$(\hat{p})$ model with $\hat{p}$ selected by AIC, first differencing the data if appropriate to remove deterministic trend effects and possible non-stationary trend unit roots. For the seasonal adjustment and the seasonal factors, we also remove the first and last three years of data. For the differenced raw data AIC yields $\hat{p} = 16$, but $\hat{p} = 2$ for the trimmed (and differenced) seasonal adjustment, and $\hat{p} = 15$ for the seasonal factors. For the over-adjustment test, we set $q = 12$. The intervals for ρ are given in Table 11, where all values of ρ between .5 and 1 (using a grid mesh of size .0001) are listed such that the p-value for the corresponding null hypothesis exceeds .01.

The results indicate that the raw data has strong seasonality present at all five of the seasonal frequencies. Next, testing the seasonal adjustment for residual seasonality indicates adequacy – no seasonality is detected. The seasonal factors appear to have captured all the seasonality that was present in the raw data, because there is failure to reject the null of persistent seasonality with very high values of ρ. Finally, the test for over-adjustment indicates there are some spectral troughs of moderate scope at all the seasonal frequencies, because anti-seasonality cannot be rejected for persistencies up to .838 (for the multiple test). This is consistent with known features of seasonal adjustment filters.

5.2. NZ Immigration

We consider an analysis of daily data described in McElroy and Jach (2019). Figure 5 displays six daily imigration series of New Zealand, covering the period September 1, 1997 through July 31, 2012. The six series are labeled as NZArr, NZDep, VisArr, VisDep, PLTArr, and PLTDep. The plots show trend and seasonal behavior, and there is also a weekly effect. There is some evidence that each of these six series may not be stationary even after trend-

Table 11. Intervals for ρ, such that the corresponding null hypothesis is not rejected at a 1% level. Rows correspond to raw data, seasonally adjusted data tested for under-adjustment, seasonal factors, and seasonally adjusted data tested for over-adjustment.

Component	$\pi/6$	$2\pi/6$	$3\pi/6$	$4\pi/6$	$5\pi/6$	All
Raw	[.994, 1)	[.994, 1)	[.996, 1)	[.994, 1)	[.989, 1)	[.994, 1)
SA (Under)	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
SF	[.997, 1)	[.997, 1)	[.997, 1)	{.999}	[.992, 1)	[.998, 1)
SA (Over)	(0,.907)	(0,.870)	(0,.913)	(0,.899)	(0,.860)	(0,.838)

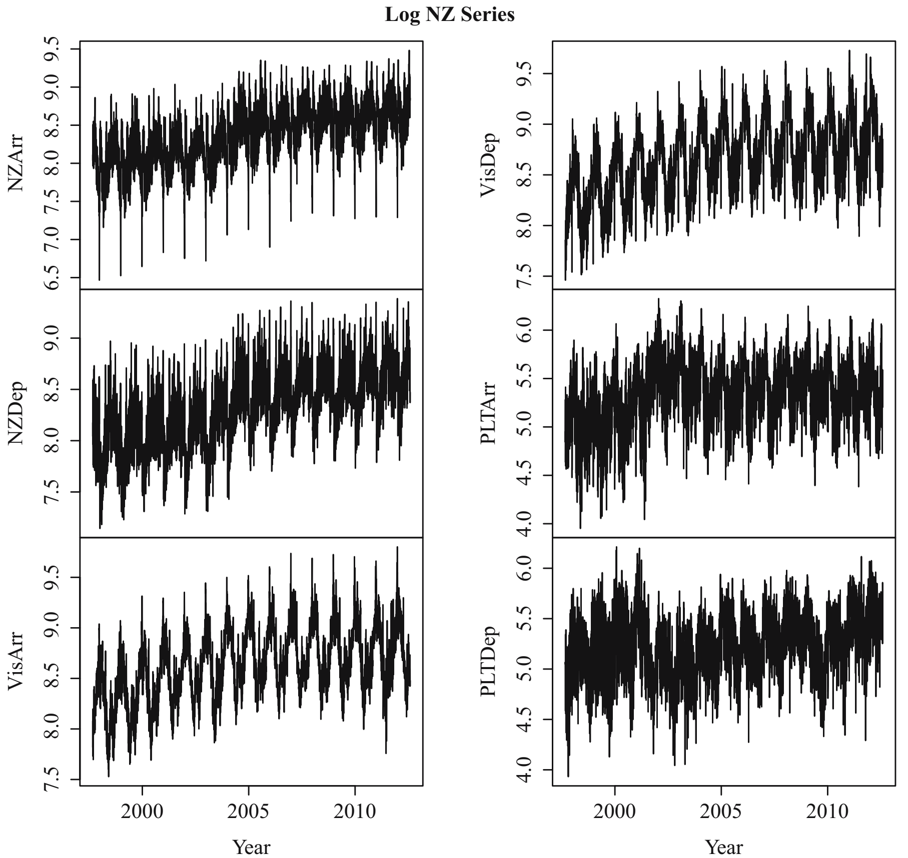

Fig. 5. Log of six New Zealand immigration series (September 1, 1997 through July 31, 2012).

differencing (there are fixed effects present), which tempers our findings accordingly. We apply the seasonality diagnostics to the differenced logged data for some of the frequencies suggested in Section 3, namely $\omega = 2\pi j/7$ for $1 \leq j \leq 3$ and $\omega = 2\pi/365.25$. An AR$(\hat{p})$ model is fitted with $\hat{p}$ selected by AIC, which is recorded in Table 12 along with the identified intervals for ρ (all values between .5 and 1 such that the p-value exceeds .01). We find that high order AR processes are needed (due to the autocorrelation present at an annual period), and that strong seasonality is present at both annual and weekly frequencies for all six series (with

Table 12. Intervals for ρ, such that the corresponding null hypothesis is not rejected at a 1% level. Rows correspond to each of the six component series.

Series	$\hat{p}$	$2\pi/365.25$	$2\pi/7$	$4\pi/7$	$6\pi/7$
NZArr	447	{.999}	{.999}	[.998, 1)	[.998, 1)
NZDep	405	{.999}	{.999}	[.998, 1)	[.998, 1)
VisArr	391	{.999}	{.999}	[.997, 1)	[.999, 1)
VisDep	404	{.999}	Ø	[.999, 1)	[.999, 1)
PLTArr	391	[.999, 1)	{.999}	[.999, 1)	[.999, 1)
PLTDep	398	[.999, 1)	[.999, 1)	[.999, 1)	[.999, 1)

the exception of the first weekly seasonal for the VisDep series) – this confirms the exploratory analysis of these series given in McElroy and Jach (2019).

5.3. Gross Domestic Product (GDP)

There has been an ongoing public debate regarding the presence of residual seasonality in GDP, which is published by the Bureau of Economic Analysis (BEA). In recent years GDP (and some of its major components) has been observed to grow at a lower rate in the first quarter (see Furman 2015; Gilbert et al. 2015; Stark 2015; Rudebusch et al. 2015; Groen and Russo 2015). These critiques have prompted research into seasonality diagnostics and seasonal adjustment at BEA (see the discussion in Lengerman et al. 2017). McCulla and Smith (2015) review BEA's response, and Phillips and Wang (2016), Lunsford (2017), and Wright (2018) delineate continuing difficulties.

We plot the logged quarterly data (left panel of Figure 6) together with growth rates (differences of logs) plotted by quarter (right panel of Figure 6), where the vertical dotted lines demarcate the different annual series for each of the four quarters. In more recent years, it appears that the first quarter is lower than the other quarters, and the question is whether this is significant. We applied the seasonality diagnostic to the differences of the logged data at frequency $\pi/2$, having obtained $\hat{p} = 16$ from AIC, and found that seasonality was rejected for all values of $\rho \in [.5, 1)$. Repeating the procedure, but focused on just the last 20 years of data, we found (with $\hat{p} = 2$) the same results. This is confirmed by examination of sample acvf and spectral plots – though when p is allowed to increase (corresponding to a likely over-specification) to 8, 12, or 16, a modest peak appears in the spectrum, somewhat off-shifted from frequency $\pi/2$. From this preliminary analysis, we do not find evidence of residual seasonality in GDP.

5.4. Census Data

We examined a collection of 233 monthly time series published by the U.S. Census Bureau, available from www.census.gov/retail/index.html. In particular, we study 65 time series of Retail Trade and Food Services (MRTS), 22 time series of Wholesale Trade: Sales and Inventories (MWTS), 4 time series of Manufacturers' Shipments, Inventories, and Orders (M3), 87 time series of Manufacturing and Trade Inventories and Sales (MTIS), and 55 time series of New Residential Construction (RES). All are monthly with a

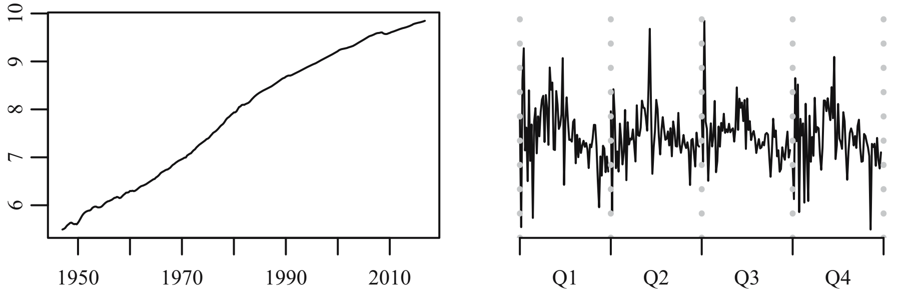

Fig. 6. Log GDP (1947 through 2016) as a time series plot (left panel) and plotted by quarter for growth rates (right panel).

start date of January 1992 or later, and with an end date of September 2019. A variety of features are present in these series: varying degrees of persistence and evolution in seasonal patterns; presence of outliers and calendrical effects; varying degrees of aggregation. In order to assess the performance of the new diagnostics, we extract seasonal adjustments and irregular components by application of the software X-13ARIMA-SEATS (U.S. Census Bureau 2015), using either the X11 option (which uses ad hoc filters in a nonlinear extraction scheme, together with ARIMA model-based forecast extension) or the SEATS option (which uses a fully model-based filtering paradigm).

We use the automatic modeling option of the software, so that calendrical effects and extremes (except for additive outliers) are removed along with the seasonal component. In many cases a log transformation is identified, and the order of differencing needed for the seasonally adjusted component is extracted from the software run; after transformation and differencing (this is not applied to the irregular), we apply the Q_s statistic as well as our proposed diagnostic for under- adjustment and over-adjusment. For the latter two diagnostics, we set the values of ρ between .98 and 1 in order to focus on the more egregious cases of residual seasonality. Setting the threshold $\alpha = .05$, any Q_s p-values above α indicate adequacy of the seasonal adjustment, whereas for the under-adjustment test we require the p-value to be less than α for adequacy. For the over- adjustment test (for which Q_s cannot be used without substantial modifications) p-values less than α indicate that over-adjustment has not occurred. We tally for each of the five batches of series the incidences of adequacy, presented in Tables 13 and 14.

The results based on the seasonal adjustments (SA) and irregular (Irr) are fairly similar, except when examining the case of over-adjustment, where there are some discrepancies (in particular, for MRTS). For both the X11 and SEATS methods, results for Q_s and the proposed test of under-adjustment are broadly similar, although Q_s is more prone to identifying residual seasonality. This is intuitive, because any degree of positive lag 12 autocorrelation will place the process in the alternative space of the Q_s test, whereas only high degrees of such positive autocorrelation will trigger the test for under-adjustment. For the test of over-adjustment most of the SA and Irr are deemed inadequate due to the large spectral troughs induced by filtering – both the X11 and SEATS methods produce this effect. Given the discussion of Bell and Hillmer (1984), these findings are not surprising.

Table 13. Number of series seasonally adjusted by X11 that are deemed to be adequate, according to whether the Q_s test, the test of under-adjustment, or the test for over-adjustment is applied, for either the seasonally adjusted (SA) component or the Irregular (Irr).

X11 Test	MRTS	MWTS	M3	MTIS	RES
SA Q_s	63	22	86	4	53
SA Under	64	22	87	4	55
SA Over	3	2	4	1	1
Irr Q_s	63	22	86	4	55
Irr Under	64	22	87	4	55
Irr Over	4	0	4	0	1
Total	65	22	87	4	55

Table 14. Number of series seasonally adjusted by SEATS that are deemed to be adequate, according to whether the Q_s test, the Root test of under-adjustment, or the Root test for over-adjustment is applied, for either the seasonally adjusted (SA) component or the irregular (Irr).

SEATS Test	MRTS	MWTS	MTIS	M3	RES
SA Q_s	64	22	85	4	52
SA Under	65	22	87	4	55
SA Over	3	1	2	0	2
Irr Q_s	65	22	86	4	55
Irr Under	65	22	87	4	55
Irr Over	6	1	2	0	2
Total	65	22	87	4	55

6. Summary

We have shown an approach to seasonality detection that aims to achieve the five criteria outlined in the Introduction. This approach depends upon a notion of seasonal persistency that relies upon the Wold Decomposition, and seasonality is measured by evaluating the corresponding AR representation at $z = \rho^{-1} e^{i\omega}$ for various $\rho \in (0, 1)$. By examining different values of ω, various periodic effects can be simultaneously investigated. For instance, in daily time series we can examine weekly effects (ω given by $2\pi/7$, $4\pi/7$, or $6\pi/7$) together with annual effects ($\omega = 2\pi/365.25$, among other integer multiples). When a series is down-sampled, or flow-aggregated, one can easily adapt the phase criteria for seasonality, although we do not mathematically derive here how the AR roots are changed by such alterations of sampling frequency.

The AR diagnostic tests are principally useful for detecting under-adjustment in a seasonal adjustment, as well as for detecting whether seasonality exists in a raw series. The former exercise is designated as testing for adequacy of seasonally adjusted data, whereas the latter is referred to as pre-testing, that is, determining whether a given series is a candidate for seasonal adjustment. (Because seasonal adjustment procedures are not idempotent in general, there is a cost associated with seasonally adjusting data that does not warrant such a procedure.) In order to detect over-adjustment, we propose the use of the MA diagnostic tests. Theory and simulation supports the use of these methods on stationary or non-stationary time series data.

The article shows how the diagnostics can be applied as a joint test over multiple frequencies, thereby allowing for a single test of seasonality (or anti-seasonality). Directionality of rejections can be determined through computing p-values as a function of the persistency ρ; by this means, we can determine whether a rejection of seasonality at a given level ρ favors the presence of more or less persistent seasonality. We stress that testing for raw seasonality avoids the difficult distribution theory of unit root tests by the device of rejecting ever more persistent formulations of stationary seasonality – this is a pragmatic approach to a thorny statistical problem.

A limitation of this article's approach is that the time series must either be stationary, or can be reduced to stationarity by a differencing polynomial. This precludes testing for fixed seasonality (given by seasonal dummies in a regression), and precludes detecting more exotic forms of seasonality (such as seasonal heteroscedasticity); mainly, the method is useful for detecting dynamic ($\rho < 1$) or unit root ($\rho = 1$) types of seasonality. A mitigating factor is that in practice an AR or ARMA model will be fitted to compute the diagnostic, and the size and power are sensitive to model selection.

7. References

Ansley, C., and W. Wecker. 1984. "Issues involved with the seasonal adjustment of economic time series, comment: on dips in the spectrum of a seasonally adjusted time series." *Journal of Business and Economics Statistics* 2(4): 323–324. DOI: https://doi.org/10.1080/07350015.1984.10509400.

Bell, W.R. 1984. "Signal extraction for nonstationary time series." *Annals of Statistics* 12(2): 646–664. DOI: https://doi.org/10.1214/aos/1176346512.

Bell, W.R., and S. Hillmer. 1984. "Issues involved with the seasonal adjustment of economic time series." *Journal of Business and Economics Statistics* 2(4): 291–320. DOI: https://doi.org/10.1080/07350015.1984.10509398.

Blakely, C., and T.S. McElroy. 2017. "Signal extraction goodness-of-fit diagnostic tests under model parameter uncertainty: formulations and empirical evaluation." *Econometric Reviews* 36(4): 447–467. DOI: https://doi.org/10.1080/07474938.2016.1140277.

Busetti, F., and A. Harvey. 2003. "Seasonality tests." *Journal of Business and Economic Statistics* 21(3): 420–436. DOI: https://doi.org/10.1198/073500103288619061.

Canova, F., and B.E. Hansen. 1995. "Are seasonal patterns constant over time? A test for seasonal stability." *Journal of Business and Economic Statistics* 13(3): 237–252. DOI: https://doi.org/10.1080/07350015.1995.10524598.

Chiu, S.T. 1988. "Weighted least squares estimators on the frequency domain for the parameters of a time series." *The Annals of Statistics* 16(3): 1315–1326. DOI: https://doi.org/10.1214/aos/1176350963.

Den Butter, F., and M. Fase. 1991. *Seasonal Adjustment as a Practical Problem*, (Vol. 1991. Amsterdam: North Holland.

Deo, R.S., and W.W. Chen. 2000. "On the integral of the squared periodogram." *Stochastic processes and their applications* 85(1): 159–176. DOI: https://doi.org/doi.org/10.1016/S0304-4149(99)00071-X.

Fase, M., J. Koning, and A. Volgenant. 1973. "An experimental look at seasonal adjustment: a comparative analysis of nine adjustment methods." *De Economist* 121: 177–180. DOI: https://doi.org/doi.org/10.1007/BF01712804.

Findley, D.F., D.P. Lytras, and T.S. McElroy. 2017. *Detecting seasonality in seasonally adjusted monthly time series*. Census Bureau research report. Available at: www.census.gov/ts/papers/rrs2017-03.pdf (accessed May 2021).

Findley, D.F., B.C. Monsell, W.R. Bell, M.C. Otto, and B.C. Chen. 1998. "New capabilities and methods of the X-12-ARIMA seasonal adjustment program." *Journal of Business and Economic Statistics* 16: 127–177. DOI: https://doi.org/doi.org/10.1080/07350015.1998.10524743.

Franses, P.H. 1994. "A multivariate approach to modeling univariate seasonal time series." *Journal of Econometrics*: 133–151. DOI: https://doi.org/doi.org/10.1016/0304-4076 (93)01563-2.

Furman, J. 2015. "Second estimate of GDP for the first quarter of 2015." *Council of Economic Advisers Blog*, May 29, 2015. Available at: https://obamawhitehouse.archives.gov/-blog/2015/05/29/second-estimate-gdp-first-quarter-2015 (accessed May 2021).

Gilbert, C.E., N.J. Morin, A.D. Paciorek, and C.R. Sahm. 2015. "Residual seasonality in GDP." *FEDS Notes*. Washington: Board of Governors of the Federal Reserve System, May 14, 2015. DOI: https://doi.org/10.17016/2380-7172.1538.

Granger, C. 1978. "Seasonality: causation, interpretation, and implications." In *Seasonal Analysis of Economic Time Series*, edited by A. Zellner: 33–46. Cambridge, MA: NBER.

Groen, J., and Russo. 2015. "The myth of first-quarter residual seasonality." *Liberty Street Economics*. June 8, 2015. Available at: https://libertystreeteconomics.newyorkfed.org/2015/06/the-myth-of-first-quarter-residual-seasonality.html (accessed May 2021).

Hylleberg, S. 1986. *Seasonality in Regression*, Orlando, Florida: Academic Press.

Hylleberg, S., R.F. Engle, C.W. Granger, and B.S. Yoo. 1990. "Seasonal integration and cointegration." *Journal of econometrics* 44(1–2): 215–238. DOI: https://doi.org/10.1016/0304-4076(90)90080-D.

Lengermann, P., N. Morin, A. Paciorek, E. Pinto, and C. Sahm. 2017. "Another Look at Residual Seasonality in GDP." *FEDS Notes*. Washington: Board of Governors of the Federal Reserve System, July 28, 2017. DOI: https://doi.org/10.17016/2380-7172.2031.

Lin, W., J. Huang, and T. McElroy. 2019. "Time series seasonal adjustment methods using regularized singular value decomposition." Published online. *Journal of Business and Economics Statistics*: 1–23. DOI: https://doi.org/10.1080/07350015.2018.1515081.

Lunsford, K.G. 2017. *Lingering residual seasonality in GDP growth*. Economic Commentary, Federal Reserve Bank of Cleveland, March 28, 2017.

Lytras, D.P., R.M. Feldpausch, and W.R. Bell. 2007. "Determining seasonality: a comparison of diagnostics from X-12-ARIMA." In Proceedings of the Third International Conference on Establishment Surveys (ICES-III June 18–21, 2007, Montreal, Canada. Available at: www.census.gov/ts/papers/ices2007dpl.pdf (accessed May 2021).

Maravall, A. 2003. "A class of diagnostics in the ARIMA-model-based decomposition of a time series." In *Seasonal Adjustment*, edited by M. Manna and R. Peronaci: pp. 23–36. Frankfurt am Main: European Central Bank.

McCulla, S.H., and S. Smith. 2015. "Preview of the 2015 annual revision of the national income and product accounts." *Survey of Current Business* 95(6): 1–8. Available at: https://apps.bea.gov/scb/pdf/2015/06%20June/0615_preview_of_2015_annual_revision_of_national_income_and_product_accounts.pdf (accessed May 2021).

McElroy, T.S. 2008. "Statistical properties of model-based signal extraction diagnostic tests." *Communications in Statistics, Theory and Methods* 37: 591–616. DOI: https://doi.org/10.1080/03610920701669785.

McElroy, T.S. 2012. "An alternative model-based seasonal adjustment that reduces over-adjustment." *Taiwan Economic Forecast and Policy* 43: 33–70. Available at: http://

www.econ.sinica.edu.tw/UpFiles/2013092817175327692/Periodicals_Pdf2013093010104847832/EC431-02.pdf (accessed May 2021).

McElroy, T., and Jach, A. 2019. "Testing collinearity of vector time series." *The Econometrics Journal* 22(2): 97–116. DOI: https://doi.org/10.1093/ectj/uty002.

McElroy, T.S., B.C. Monsell, and R.J. Hutchinson. 2018. *Modeling of holiday effects and seasonality in daily time series.* Census Bureau research report 2018-01. Available at: www.censusgov/srd/papers/pdf/RRS2018-01.pdf (accessed May 2021).

McElroy, T., and D. Politis. 2020. *Time Series: a First Course with Bootstrap Starter*. New York: Chapman and Hall.

McElroy, T.S. and A. Roy. 2017. *Detection of seasonality in the frequency domain*. Census Bureau research report, 2017–01. Available at: www.census.gov/ts/papers/rrs2017-01.pdf (accessed May 2021).

McElroy, T., and A. Roy. 2018. "The inverse Kullback Leibler method for fitting vector moving averages." *Journal of Time Series Analysis* 39: 172–191. DOI: https://doi.org/10.1111/jtsa.12276.

Moulton, B.R., and B.D. Cowan. 2016. "Residual seasonality in GDP and GDI: findings and next steps." *Survey of Current Business* 96(7): 1–6. Available at: https://apps.bea.gov/scb/pdf/2016/07%20July/0716_residual_seasonality_in_gdp_and_gdi.pdf (accessed May 2021).

Nerlove, M. 1964. "Spectral analysis of seasonal adjustment procedures." *Econometrica* 32: 241–286. DOI: https://doi.org/10.2307/1913037.

Phillips, K., and J. Wang. 2016. "Residual seasonality in U.S. GDP data." FRB of Dallas Working Paper No. 1608. DOI: https://doi.org/10.24149/wp1608.

Proietti, T. 1996. "Persistence of shocks on seasonal processes." *Journal of Applied Econometrics* 11(4): 383–398. DOI: https://doi.org/10.1002/(SICI)1099-1255(199607)11:4¡383:AID-JAE403¿3.0.CO;2-3.

Rudebusch, G.D., D. Wilson, and T. Mahedy. 2015. "The puzzle of weak first-quarter GDP growth." *FRBSF Economic Letter*, May 18, 2015. Available at: https://www.frbsf.org/economic-research/publications/economic-letter/2015/may/weak-first-quarter-gdp-residual-seasonality-adjustment/ (accessed May 2021).

Sims, C. 1978. "Comments on 'Seasonality: causation, interpretation, and implications' by Clive W.J. Granger." In *Seasonal Analysis of Economic Time Series*, edited by A. Zellner: 47–49. Cambridge, MA: NBER.

Soukup, R., and D. Findley. 1999. "On the spectrum diagnostics used by X-12-ARIMA to indicate the presence of trading day effects after modeling or adjustment." In 1999 Proceedings American Statistical Association: Alexandria. August 8–12, 1999, Baltimore, Maryland, USA. Available at: www.census.gov/ts/papers/rr9903s.pdf (accessed May 2021).

Stark, T. 2015. "First quarter in the national income and product accounts." *Research Rap*, Federal Reserve Bank of Philadelphia, May 14, 2015. Available at: https://www.philadelphiafed.org/-/media/frbp/assets/economy/reports/research-rap/2015/first_quarters_national_income_product_accounts.pdf (accessed May 2021).

Tanaka, K. 1996. *Time series analysis: nonstationary and noninvertible distribution theory*. New York: John Wiley and Sons.

Tukey, J. 1978. "Comments on 'Seasonality: causation, interpretation, and implications' by Clive W.J. Granger." In *Seasonal Analysis of Economic Time Series*, edited by A. Zellner: 50–54. Cambridge, MA: NBER.

U.S Census Bureau. 2015. *X-13ARIMA-SEATS reference manual*. U.S. Census Bureau, Washington D.C., USA. Available at: www.census.gov/ts/x13as/docX13AS.pdf (accessed date May 2021).

Wright, J.H. 2018. "Seasonal adjustment of NIPA data." *NBER working paper*, 24895, August 2018. DOI: https://doi.org/10.3386/w24895.

Received May 2019
Revised June 2020
Accepted September 2020