A Seasonality Diagnostic Based Upon Multi-Step Ahead Forecasting Errors

Tucker McElroy[1]
U.S. Census Bureau

Seasonal Adjustment Practitioners Workshop
September 11, 2024

[1]This presentation 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. All time series analyzed in this presentation are from public or external data sources.

Outline

- Seasonality diagnostics motivation
- Heuristics: seasonal persistence and intra-seasonal association
- Seasonality measures defined
- Illustrations

Overview

Seasonality is present in many regularly-spaced economic time series

Statistical agencies seek to identify and remove (seasonally adjust) the seasonality, so as to allow easier visualization of trends and cycles

Challenge: how to define (or measure) seasonality? What diagnostic corresponds to our definition?

Available diagnostics

There are many diagnostics available, each based on a certain definition of seasonality:

- Model-based F (MBF): test of seasonal means in a linear model with time series errors

- QS: seasonal lag autocorrelation

- Visual Significance (VS): peak in spectral density at seasonal frequency

- Root: seasonal roots in autoregressive polynomial

Critique

1. These diagnostics cannot be defined without first postulating a form of process: RegARIMA for MBF, difference stationary for QS and VS, autoregressive for Root.

2. These diagnostics assess different types of seasonality: fixed for MBF, dynamic for QS and VS and Root.

3. These diagnostics may classify non-seasonal processes as seasonal: e.g., QS.

4. These diagnostics are not easily conveyed to non-experts.

Goal

Define a new seasonality measure (and diagnostic) with these features:

- Intuitive definition that is broad, such that when specialized to familiar processes (e.g., difference stationary, RegARIMA, autoregressive) they have expected behavior on seasonal and non-seasonal processes.
- Avoids classification problems by incorporating both **seasonal persistence** and **intra-seasonal association**.
- Should be scale-free.

Problem with seasonal autocorrelation

Let s be the integer seasonal period.

Intuitively, a high association between observations s lags apart is necessary to describe seasonality.

For a stationary process, this would be measured with the lag s autocorrelation.

But: for the AR(1) process, which is non-seasonal, the lag s autocorrelation is ϕ^s. For $s = 4$ and $\phi = .95$, this yields $.815$, a high value – a mis-classification.

Modifying seasonal autocorrelation

The AR(1) case is instructive: the seasonal autocorrelation is really driven by the high lag 1 autocorrelation, as season-to-season there is a high association.

We should *condition* on this **intra-seasonal association**, to balance seasonal persistence.

Consider the seasonal sub-series (i.e., annual time series for each season); if these have a similar pattern, then we say there is intra-seasonal association.

Seasonal persistence

Let $\{X_t\}$ be our time series. We first define **seasonal persistence** by modifying the seasonal autocovariance slightly: we condition on the recent past (denoted by $\{X_{t:}\}$), which isolates current observations. The seasonal persistence is defined as

$$\Xi_s = \text{Cov}[X_{t+s+1}, X_{t+1} | X_{t:}]. \tag{1}$$

The conditioning removes time-dependence from the measure for many processes, since (1) can be expressed as the covariance of $s+1$-step ahead and 1-step ahead forecast errors.

Intra-seasonal association

Intra-seasonal association is measured by the s-step ahead forecast error

$$X_{t+s} - \widehat{X}_{t+s|t:},$$

since this will tend to be large when the seasonal sub-series are tightly linked.

But intra-seasonal association can be high whether or not the process has seasonality.

If the process is seasonal, intra-seasonal association should be low.

Seasonal measures

We can account for intra-seasonal association in the seasonal persistence measure by introducing conditioning upon the s-step ahead forecast error, or (equivalently)

$$\Omega_s = \mathrm{Cov}[X_{t+s+1}, X_{t+1} | X_{t+s}, X_{t:}].$$

It is straightforward to show that

$$\Omega_s = \Xi_s - \mathrm{Cov}[X_{t+s+1}, X_{t+s} | X_{t:}] \, \mathrm{Var}[X_{t+s} | X_{t:}]^{-1} \, \mathrm{Cov}[X_{t+s}, X_{t+1} | X_{t:}]$$

Normalized form

Because Ω_s is not scale invariant, we can also consider the partial correlation:

$$\begin{aligned} \Upsilon_s &= \text{Corr}[X_{t+s+1}, X_{t+1} | X_{t+s}, X_{t:}] \\ &= \frac{\Omega_s}{\sqrt{\text{Var}[X_{t+s+1} | X_{t+s}, X_{t:}] \, \text{Var}[X_{t+1} | X_{t+s}, X_{t:}]}}. \end{aligned}$$

This gives a seasonality measure with values in $[-1, 1]$, with 0 indicating no seasonality, and positive values indicating moderate to high degrees of seasonality. Negative values are designated as *anti-seasonality*.

Stationary forecast errors

For many classes of processes the forecast errors will be stationary (jointly across leads), and hence

$$\upsilon_h^{(k)} = \mathsf{Cov}[X_{t+k}, X_{t+h+k} | X_{t:}]$$

does not depend on t. Then

$$\Omega_s = \upsilon_s^{(1)} - \upsilon_{s-1}^{(1)} \upsilon_1^{(s)} / \upsilon_0^{(s)}$$

$$\Upsilon_s = \frac{\upsilon_s^{(1)} - \upsilon_{s-1}^{(1)} \upsilon_1^{(s)} / \upsilon_0^{(s)}}{\sqrt{(\upsilon_0^{(1)} - \upsilon_{s-1}^{(1)}{}^2 / \upsilon_0^{(s)})(\upsilon_0^{(s+1)} - \upsilon_1^{(s)}{}^2 / \upsilon_0^{(s)})}}.$$

Difference stationary processes

We say $\{X_t\}$ is difference stationary if there exists a unit root polynomial $\delta(z)$ such that $\delta(B)X_t = W_t$ is stationary, say with MA (∞) representation $W_t = \psi(B)Z_t$, with $\{Z_t\}$ white noise of variance σ^2 and $\psi(z)$ causal. (B is the backshift operator.) Then forecast errors are stationary, and

$$\upsilon_h^{(k)} = \sigma^2 \sum_{\ell=0}^{k-1} \xi_\ell \xi_{\ell+h},$$

where $\xi(z) = \psi(z)/\delta(z)$. Also

$$\Omega_s = \sigma^2 \left(\xi_s - \xi_{s-1} \sum_{\ell=0}^{s-1} \xi_\ell \xi_{\ell+1} / \sum_{\ell=0}^{s-1} \xi_\ell^2 \right).$$

AR(1) process

$\{X_t\}$ is an AR(1), so that $\psi(z) = (1 - \phi z)^{-1}$ and $\xi_j = \psi_j = \phi^j$:

$$\Xi_s = \sigma^2 \phi^s$$
$$\Omega_s = \sigma^2 (\phi^s - \phi^{s-1}\phi) = 0.$$

So there is no seasonality present once the intra-seasonal association is accounted for.

United States™ Census Bureau

Cyclic AR(2) process

$\{X_t\}$ is an AR(2) with complex conjugate seasonal roots: $\psi(z) = (1 - 2\rho\cos(\omega)z + \rho^2 z^2)^{-1}$, which corresponds to autoregressive roots $\rho^{-1}\exp\{\pm i\omega\}$ (for $0 < \rho < 1$), where $\omega = 2\pi\ell/s$ for some integer ℓ. Then $\xi_j = \psi_j = \rho^j \cos(\omega j)$, and

$$\Xi_s = \sigma^2 \rho^s$$

$$\Omega_s = \sigma^2 \rho^s \left(1 - \cos(\omega) \sum_{\ell=0}^{s-1} \rho^{2\ell} [\cos(\omega(2\ell+1)) + \cos(\omega)] / \sum_{\ell=0}^{s-1} \rho^{2\ell} [\cos(\omega(2\ell)) + 1] \right)$$

Setting $s = 4$, $\Upsilon_4 = .326$ for $\rho = .8$, and $\Upsilon_4 = .454$ for $\rho = .9$. Therefore, such a process exhibits seasonality.

SAR(1) process

$\{X_t\}$ is a seasonal autoregression of order 1 (or SAR(1)), so that $\psi(z) = (1 - \phi_s z^s)^{-1}$. So $\xi_j = \psi_j$ is zero unless $j = sk$ for integer k, in which case $\xi_j = \phi_s^k$. Then

$$\Xi_s = \sigma^2 \phi_s$$
$$\Omega_s = \sigma^2 \phi_s$$
$$\Upsilon_s = \phi_s / \sqrt{1 + \phi_s^2}.$$

The maximum value of Υ_s for this process is $1/\sqrt{2}$.

Seasonal difference process

$\{X_t\}$ is a SARIMA process with differencing polynomial $\delta(z) = 1 - z^s$, so that $\xi(z) = \psi(z) + z^s \psi(z) + \ldots$. Then

$$\Xi_s = \sigma^2 (\psi_s + 1)$$

$$\Omega_s = \sigma^2 \left(1 + \psi_s - \psi_{s-1} (\psi_{s-1} + \sum_{\ell=0}^{s-1} \psi_\ell \psi_{\ell+1}) / \sum_{\ell=0}^{s-1} \psi_\ell^2 \right).$$

Seasonal difference process

Further, if $\psi(z) = 1 - \theta_s z^s$, then

$$\Xi_s = \Omega_s = \sigma^2 (1 - \theta_s)$$

$$\Upsilon_s = (1 - \theta_s) / \sqrt{1 + (1 - \theta_s)^2}.$$

As θ_s approaches 1, the nonstationary process resembles a white noise (due to cancellation of operators), and the seasonality measure tends to zero. If instead $\theta_s \rightarrow -1$, then $\Upsilon_s \rightarrow 2/\sqrt{5}$, which is a higher value than that attainable by the SAR(1).

Seasonal and regular difference process

$\{X_t\}$ is a SARIMA process with differencing polynomial $\delta(z) = (1 - z^s)(1 - z)$. Letting $U(z) = 1 + z + \ldots + z^{s-1}$, we find that $\xi(z) = \psi(z)U(z) + 2z^s\psi(z)U(z) + \ldots$. It follows that

$$\Xi_s = \sigma^2(2 + \sum_{k=1}^{s} \psi_k).$$

Airline process

For the airline process $\psi(z) = (1 - \theta_1 z)(1 - \theta_s z^s)$, so that

$$\Xi_s = \sigma^2 (2 - \theta_1 - \theta_s)$$

$$\Omega_s = \sigma^2 \left((2 - \theta_1 - \theta_s) \left[\frac{1 + (s-2)(1-\theta_1)^2}{1 + (s-1)(1-\theta_1)^2} \right] - (1-\theta_1)^2 \left[\frac{1 + (s-2)(1-\theta_1)}{1 + (s-1)(1-\theta_1)^2} \right] \right).$$

So high positive values of θ_s lessen the seasonal persistence. Also, positive values of θ_1 tend to increase seasonality, up to a point, by decreasing the intra-seasonal association. See Figures 1 and 2 for the case of $s = 4$.

Airline process

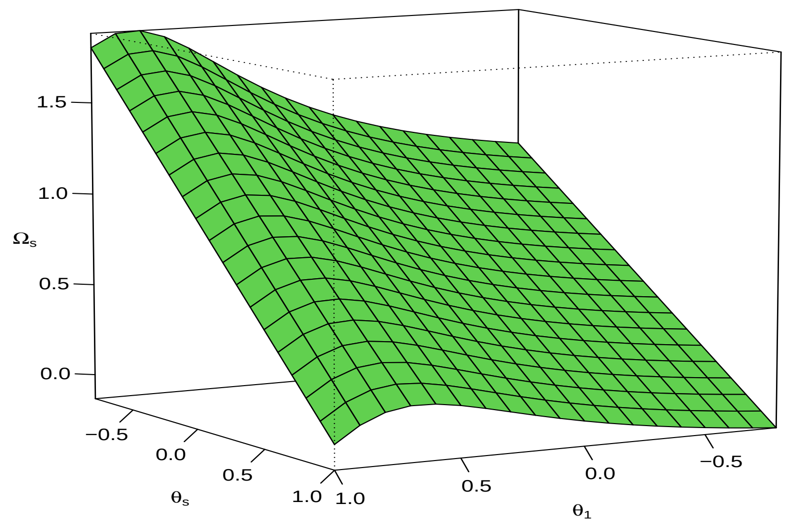

Figure 1: Values of Ω_4 for an airline process, as a function of θ_1 and θ_s.

Airline process

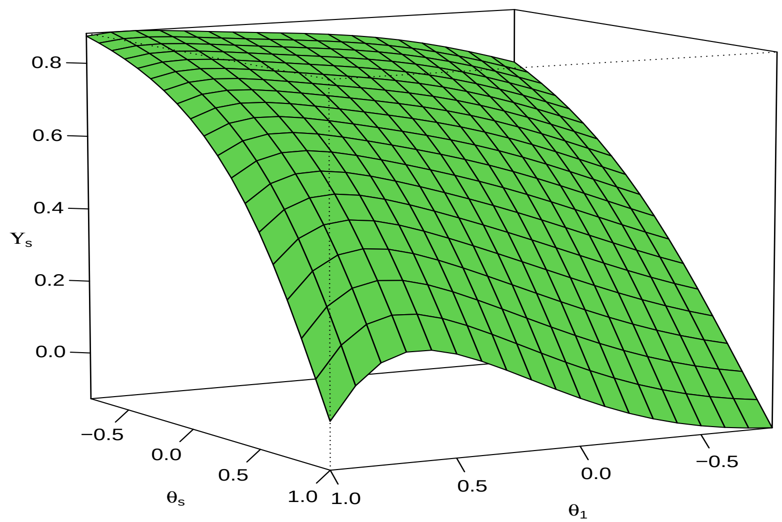

Figure 2: Values of Υ_4 for an airline process, as a function of θ_1 and θ_s.

Seasonal Roots

We can analyze ARIMA processes via the roots of its autoegressive polynomial:

- Seasonal persistency Ξ_s is driven by AR roots with near-unit magnitude and seasonal frequency (argument is a multiple of $2\pi/s$)

- Intra-seasonal association is high when a reciprocal root is strong and close to the lag one autocorrelation

Other Diagnostics

The seasonality measures can be connected to VS, Root, and QS.

- Spectral peaks occur at frequencies associated with the argument of AR roots with modulus close to one. So Ω_s is connected to Root and VS diagnostics.

- For ARIMA processes a large value of Ω_s indicates a large lag s autocorrelation, so the QS measure will be large; conversely, QS can be large due to high intra-seasonal association (e.g., AR(1)) even when Ω_s is small.

Inference

We can estimate Ω_s and Υ_s by replacing covariances of forecast errors by sample covariances of such, computed in-sample.

A CLT for both estimators has been derived, but the asymptotic variance depends on quantities not determined by the null hypothesis.

Testing: we may wish to test the null hypothesis $\Omega_s = 0$, which is equivalent to $\upsilon_s^{(1)} \upsilon_0^{(s)} - \upsilon_{s-1}^{(1)} \upsilon_1^{(s)} = 0$.

Testing

Consider the estimator

$$\widehat{\theta}_T = \widehat{\upsilon}_s^{(1)} \widehat{\upsilon}_0^{(s)} - \widehat{\upsilon}_{s-1}^{(1)} \widehat{\upsilon}_1^{(s)},$$

which tends to zero if $\Omega_s = 0$. This has a CLT with unknown variance, so we propose studentizing by some S_T (based on partial sums of forecast error cross-products) and obtain ($\{B_r\}$ is standard Brownian Motion)

$$T^{1/2} \frac{\widehat{\theta}_T - \mu_Z}{\sqrt{S_T}} \stackrel{\mathcal{L}}{\Longrightarrow} \frac{2B_1}{\sqrt{\int_0^1 (B_r - rB_1)^2 dr}}.$$

Numerical Results

We simulate 10^4 draws of a quarterly SAR(1) process with $\phi_s = .8$ and unit innovation variance, for which $\Omega_s = .8$.

To illustrate the size properties, we center the statistic by $\mu_Z = .8$ (first column), and for power properties we set $\mu_Z = 0$ (second column).

Numerical Results

Table 1: Upper one-sided rejection rates for studentized seasonal measure test statistic applied to a SAR(1) with $\phi_s = .8$, by number of years n, nominal level α, and centering by μ_Z.

n	α	$\mu_Z = .8$	$\mu_Z = 0$
	.01	.005	.541
$n = 20$	.05	.031	.883
	.10	.081	.976

Summary

- Seasonality is defined as seasonal persistence that is not explained by intra-seasonal association

- Seasonal persistence is defined as lag s covariance conditional on the past

- Intra-seasonal association is defined as lead s forecast error

- Ξ_s measures seasonal persistence, Ω_s measures seasonal persistence conditional on intra-seasonal association, and Υ_s is a normalized Ω_s

- These measures have desired values on various SARIMA processes; correct classification

Future work

- Investigate seasonality measures on seasonally heteroscedastic processes and stochastic seasonal means (including RegARIMA) processes

- Explore the impact of seasonal adjustment on seasonality measures

- Empirical testing

Contact

tucker.s.mcelroy@census.gov