A Diagnostic for Seasonality Based Upon Autoregressive Roots

Tucker McElroy (U.S. Census Bureau)

2018 Seasonal Adjustment Practitioners Workshop
April 26, 2018

Disclaimer

This presentation 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 necessarily those of the U.S. Census Bureau.

Outline

- Motivation and Desiderata for Seasonality Diagnostics
- Critique of MB-F, VS, and QS
- Foundations for the AR Root Diagnostic
- Statistical Theory
- Application to Furniture Retail Series
- Conclusions and Extensions

Motivation

- Recent critiques of GDP seasonal adjustment (SA): elements of the public greatly care about quality of SA in high-profile time series
- Agencies (including Census) are moving towards publication/application of weekly and daily time series, which have multiple types of seasonality present at non-integer periods
- Seasonal heteroscedasticity invalidates current diagnostics based upon stationarity assumptions

Desiderata

Seasonality diagnostic wishlist:

1. Rigorous statistical theory
2. Diagnostic is necessary and sufficient to capture seasonality
3. Applicable to diverse sampling frequencies (e.g., quarterly, monthly, weekly, daily, etc.)
4. Identifies non-integer period effects
5. Addresses over-adjustment as well as under-adjustment

Proposed Verbal Definition

Seasonality: persistency in a time series over seasonal periods that is not explainable by intervening time periods

Quarterly Parsing: persistency in a quarterly time series from year to year that is not explainable by inter-quarterly dynamics

Critique of Model-Based F Test

MB-F Scorecard:

1. Has rigorous statistical theory (though, depends upon a parametric model)
2. Only captures deterministic (fixed) seasonality, is not effective at detection of dynamic (time-varying) seasonality
3. Can be adapted to diverse sampling frequencies (additional regressors)
4. Can identify non-integer periods
5. Does not assess negative seasonal lag correlation (a symptom of over-adjustment)

Critique of Visual Significance

VS Scorecard:

1. Has rigorous statistical theory (for the case of a tapered spectral estimator)
2. Seasonality can be present without a peak in the spectrum being manifested due to the superposition principle; VS is not necessary
3. Can be adapted to diverse sampling frequencies (frequency domain)
4. Can identify non-integer periods (frequency domain)
5. Could be adapted to detect seasonal spectral troughs (a symptom of over-adjustment)

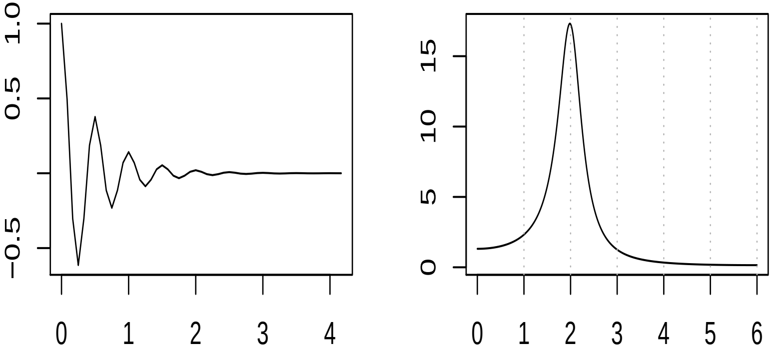

Figure: Autocorrelation function (left panel) and spectral density (right panel) for a seasonal AR(2) process

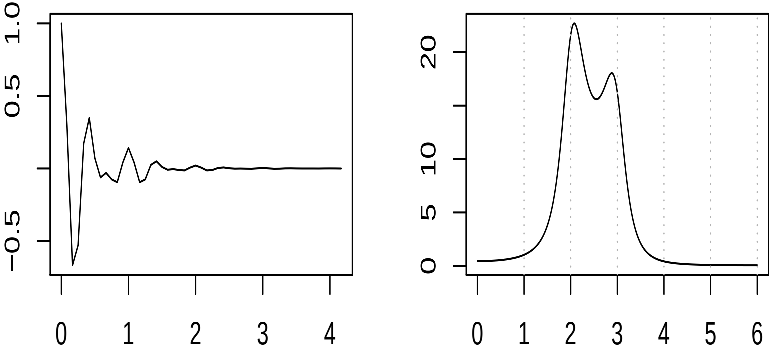

Figure: Autocorrelation function (left panel) and spectral density (right panel) for a seasonal AR(4) process

QS Scorecard:

1. Ad hoc asymptotic theory (based on simulations)
2. Diagnostic can flag non-seasonal processes as seasonal; QS is not necessary
3. Can be adapted to diverse sampling frequencies (seasonal lags of autocorrelations)
4. Cannot identify non-integer periods
5. Could be adapted to assess negative seasonal lag correlation (a symptom of over-adjustment)

Critique of QS Test

Counter-Example: a (non-seasonal) AR(1) with parameter ϕ has autocorrelation function $\rho_h = \phi^h$, which for $h = s$ (the seasonal period) is high if ϕ is large, falsely indicating seasonality according to QS

A Fatal Simulation: with $\phi = .98$ and $s = 4$ we obtain $\rho_4 = .92$. Consider 10^5 simulations of a 20-year sample of this process: the empirical type I error rate is .975 based on the nominal of .05, i.e., a 97.5% chance of falsely indicating seasonality!!!

QS fails: because it has left off necessity, i.e. the second part of our definition – we need to screen out cases where seasonal lag autocorrelation is high due mainly to intra-seasonal effects

Fixing QS: look for oscillatory patterns in the autocorrelation function (acf), i.e., peaks at seasonal lags and lower values nearby. We need a nice representation...

Hints from VS: spectral plots can be deceiving, but they are based on autoregressive (AR) root structure...

Fact: if the spectral density

$$f(\lambda) = \sum_{h \in \mathbb{Z}} \gamma_h \, e^{-ih\lambda}$$

is positive, where γ_h is the autocovariance function (acvf), then f is dense in the space of AR spectra, i.e., there exists an AR(p) approximation for p sufficiently large

Foundation: the AR Spectrum

Assuming the roots ζ_j are distinct, write

$$\phi(z) = 1 - \sum_{j=1}^{p} \phi_j \, z^j = \prod_{j=1}^{p} (1 - \zeta_j^{-1} \, z)$$

for the AR polynomial, so that approximately (σ^2 is prediction error variance)

$$f(\lambda) = \sigma^2 \left| \phi(e^{-i\lambda}) \right|^{-2}$$

Foundation: the AR Acf

By calculus of residues, for $h \geq 0$

$$\gamma_h = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda)\, e^{ih\lambda}\, d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\sigma^2}{|\phi(e^{-i\lambda})|^2}\, e^{ih\lambda}\, d\lambda$$

$$= \frac{\sigma^2}{2\pi i} \int_{\partial D} \frac{z^{h-1}}{\phi(z)\,\phi(z^{-1})}\, dz = \sigma^2 \sum_{j=1}^{p} \frac{-\zeta_j^{-h}}{\phi(\zeta_j^{-1})\,\zeta_j\,\nabla\phi(\zeta_j)}.$$

Hence the acf is

$$\rho_h = \sum_{j=1}^{p} c_j\, \zeta_j^{-h} = \sum_{j=1}^{p} \frac{\left(\phi(\zeta_j^{-1})\,\zeta_j\,\nabla\phi(\zeta_j)\right)^{-1}}{\sum_{k=1}^{p} \left(\phi(\zeta_k^{-1})\,\zeta_k\,\nabla\phi(\zeta_k)\right)^{-1}}\, \zeta_j^{-h}$$

Polar: each root ζ_j has a polar decomposition:
$\zeta = |\zeta| \exp\{i \arg(\zeta)\}$ in terms of magnitude $|\zeta|$ and phase $\arg(\zeta) = \arctan(\Im\zeta / \Re\zeta)$.

$$\zeta_j^{-h} = (1/|\zeta_j|)^h \exp\{-i\, h \arg(\zeta_j)\}$$

Summary: the acf is expressed as a linear combination of damped exponentials ζ_j^{-h} (reciprocal modulus is damping, phase yields periodicity of sinusoids). For roots close to unity, the coefficients are close to maximal, and the corresponding oscillatory effects will be evident in the acf (and acvf)

Foundation: Extensions to Over-Adjustment

Over-adjustment: seasonal adjustment filters that smooth too much extract non-seasonal dynamics and place them in the seasonal, resulting in troughs in the spectrum at seasonal frequencies

MA Roots and Inverse Autocorrelations: spectral troughs are peaks in the reciprocal spectrum. These can be assessed through moving average (MA) roots, using an MA sieve; the inverse autocorrelation function (iacf) can then be decomposed in terms of MA root magnitude and phase

Persistency and Anti-persistency: large magnitude AR roots yield persistent dynamics (high autocorrelation at lags corresponding to reciprocal phase), whereas large magnitude MA roots yield anti-persistent dynamics (high inverse autocorrelation at lags corresponding to reciprocal phase)

Statistical Theory: AR Coefficients

If the process is stationary (we can trend-difference first), and we estimate the AR(p) via OLS (with p selected by AIC), then

$$\sqrt{T}\,(\underline{\widehat{\phi}} - \underline{\phi}) \stackrel{\mathcal{L}}{\Longrightarrow} \underline{Z} \sim \mathcal{N}(0, V)$$

for $\underline{\phi} = [\phi_1, \ldots, \phi_p]'$ and $V = \sigma^2\,\Gamma_p^{-1}$, σ^2 is the prediction error variance, and Γ_p is the $p \times p$ dimensional Toeplitz covariance matrix of the process

Statistical Theory: AR Roots

Using Rouché's Theorem and Delta method, we obtain

$$A = \begin{bmatrix} \frac{\Re\widetilde{\zeta}}{|\widetilde{\zeta}|} \Re[\nabla\phi(\widetilde{\zeta})]^{-1} \underline{\widetilde{\zeta}}' + \frac{\Im\widetilde{\zeta}}{|\widetilde{\zeta}|} \Im[\nabla\phi(\widetilde{\zeta})]^{-1} \underline{\widetilde{\zeta}}' \\ \frac{\Re\widetilde{\zeta}}{|\widetilde{\zeta}|^2} \Im[\nabla\phi(\widetilde{\zeta})]^{-1} \underline{\widetilde{\zeta}}' - \frac{\Im\widetilde{\zeta}}{|\widetilde{\zeta}|^2} \Re[\nabla\phi(\widetilde{\zeta})]^{-1} \underline{\widetilde{\zeta}}' \end{bmatrix}$$

$$\sqrt{T}\left(|\widehat{\zeta}| - |\widetilde{\zeta}|, \arg(\widehat{\zeta}) - \arg(\widetilde{\zeta})\right) \stackrel{\mathcal{L}}{\Longrightarrow} A\,\underline{Z}.$$

All the quantities in A can be consistently estimated by plugging in $\widehat{\zeta}$ for $\widetilde{\zeta}$, and $\widehat{\phi}$ for ϕ

Statistical Theory: Null Hypothesis for Seasonality

We say that a seasonal root generates δ-seasonality if and only if $|\zeta| \leq 1 + \delta$, where $\delta \leq .1$ is suggested by numerical work. So for some $1 \leq j \leq \lfloor s/2 \rfloor$

$$H_0 : |\zeta| = 1 + \delta \text{ and } \arg(\zeta) = \pm \pi j / s$$
$$H_a : |\zeta| > 1 + \delta \text{ or } \arg(\zeta) \neq \pm \pi j / s$$

So the null states that ζ is a seasonal root that generates δ-seasonality

Statistical Theory: Wald Test Statistic for Seasonality

We test this null with a Wald statistic, where we evaluate ζ_0 at the boundary null root:

$$\mathcal{S}(\zeta_0) = T \left[\begin{array}{c} |\widehat{\zeta}| - |\zeta_0| \\ \arg(\widehat{\zeta}) - \arg(\zeta_0) \end{array} \right]' \left[A\, V\, A' \right]^{-1} \left[\begin{array}{c} |\widehat{\zeta}| - |\zeta_0| \\ \arg(\widehat{\zeta}) - \arg(\zeta_0) \end{array} \right]$$

$$\stackrel{\mathcal{L}}{\Longrightarrow} \chi^2_2$$

Application to Furniture Retail Series: Goals

Data: series 442 (Furniture and Home Furnishings Stores) of the Advance Monthly Retail Trade Report, January 1992 through December 2012

Goals: test the raw series for seasonality ($\delta = 0$), verify seasonality in the seasonal factors ($\delta = 0$), and test the adjusted series (performed by X-12-ARIMA) for residual seasonality ($\delta = .1$)

Application to Furniture Retail Series: Display

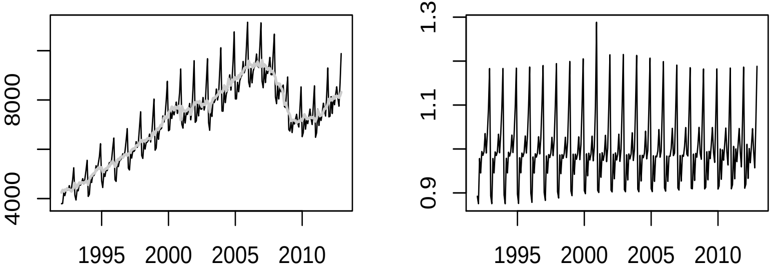

Figure: Retail series 442 (Furniture and Home Furnishings Stores), with seasonal adjustment (grey, left panel) and seasonal factors (right panel)

Application to Furniture Retail Series: Analysis of Raw

Summary of results:

- Fitted an AR(p) model to the differenced logged data, obtaining $\widehat{p} = 23$
- Obtained eleven seasonal roots (of unity) and twelve non-seasonal root. All seasonal roots fail to reject $\delta = 0$ null, all other roots do reject null
- Econometric bonus: roots 12 and 13 correspond to a period of roughly 6.08 years, and likely corresponds to a business cycle

Application to Furniture Retail Series: Results, Part I

Raw Data			X12 Seasonal Factors		
$\arg(\widehat{\zeta})s/(2\pi)$	$\lvert\widehat{\zeta}\rvert$	$\mathcal{S}(\zeta)$ p-value	$\arg(\widehat{\zeta})s/(2\pi)$	$\lvert\widehat{\zeta}\rvert$	$\mathcal{S}(\zeta)$ p-value
−4.00080	0.99902	1.0000	−4.00152	0.99912	1.0000
+4.00080	0.99902	1.0000	+4.00152	0.99912	1.0000
−2.99944	0.99999	1.0000	−2.99911	0.99976	1.0000
+2.99944	0.99999	1.00000	+2.99911	0.99976	1.0000
−5.00209	1.00023	0.50283	−5.00170	1.00030	0.6927
+5.00209	1.00023	0.50283	+5.00170	1.00030	0.6927
−1.99497	1.00359	0.52510	−1.99574	1.00174	0.56128
+1.99497	1.00359	0.52510	+1.99574	1.00174	0.56128
−0.99663	1.00527	0.51773	−0.99868	1.00174	0.77719
+0.99663	1.00527	0.51773	+0.99868	1.00174	0.77719
−6.00000	1.01213	0.58792	−6.00000	1.00580	0.6979
−0.16451	1.05488	0.00000	0.00000	1.11313	0.0000

Application to Furniture Retail Series: Results, Part II

Raw Data			X12 Seasonal Adjustment		
$\arg(\widehat{\zeta})s/(2\pi)$	$\lvert\widehat{\zeta}\rvert$	$\mathcal{S}(\zeta)$ p-value	$\arg(\widehat{\zeta})s/(2\pi)$	$\lvert\widehat{\zeta}\rvert$	$\mathcal{S}(\zeta)$ p-value
+0.16451	1.05488	0.00000	0.00000	1.07755	0.0000
−6.00000	1.31689	0.81077	−5.43569	1.20062	0.0000
−6.00000	1.64682	0.87328	+5.43569	1.20062	0.0000
−3.45904	1.10981	0.00000	−3.49204	1.19769	0.00090
+3.45904	1.10981	0.00000	+3.49204	1.19769	0.00090
−2.29998	1.11912	0.00000	−2.28444	1.15781	0.00093
+2.29998	1.11912	0.00000	+2.28444	1.15781	0.00093
−1.20483	1.23660	0.00046	−1.21102	1.34944	0.04321
+1.20483	1.23660	0.00046	+1.21102	1.34944	0.04321
−4.59082	1.23726	0.00024	−4.22049	1.34161	0.19874
+4.59082	1.23726	0.00024	+4.22049	1.34161	0.19874

Application to Furniture Retail Series: Analysis of Seasonal Factors

Summary of results:

- Fitted an AR(p) model to the logged seasonal factors, obtaining $\widehat{p} = 12$
- Obtained eleven seasonal roots (of unity) and one non-seasonal root. All seasonal roots fail to reject $\delta = 0$ null, the other root does reject null
- Concern: why is a non-seasonal root in there? Over-adjustment?

Application to Furniture Retail Series: Analysis of Seasonal Adjustment

Summary of results:

- Fitted an AR(p) model to the differenced logged seasonal adjustment, obtaining $\widehat{p} = 11$
- Obtained eleven roots, none of which appears to have seasonal phase
- Potential concern: one pair of roots, near to the fourth seasonal frequency, have p-value of .199. (If we alter δ to zero, the p-value drops to .083, hence only dynamic seasonality is present.)
- Econometrician's concern: where's my business cycle!?

Conclusions

Key Takeaways:

- AR roots can capture persistencies in the data
- AR roots form the basis of a seasonality diagnostic satisfying five desiderata...
- Very fast and simple implementation (a few dozen lines in R, fit with OLS)
- Extensions to seasonal heteroscedastic and multivariate time series possible (under development)...

Critique of AR Roots Seasonality Test

AR Roots Scorecard:

1. Has rigorous statistical theory (semi-parametric through AR sieve)
2. Captures deterministic (fixed) seasonality ($\delta = 0$) and dynamic (time-varying) seasonality ($\delta > 0$)
3. Can be adapted to diverse sampling frequencies, and handle multiple seasonalities (e.g., weekly and annual periods in daily data) through phase
4. Can identify non-integer periods (e.g., monthly period in daily data) through phase
5. Can assess over-adjustment via anti-persistency test through MA roots

Extensions

Business cycle analysis: look for phase of AR roots corresponding to periods between 2 and 10 years, and test with a composite null

Adequate seasonal adjustment: we can design a concurrent filter based on a partial fraction decomposition of $\phi(z)$ into its seasonal and non-seasonal roots, and obtain seasonal factors and seasonal adjustments that are guaranteed to be free of under- and over-adjustment problems

Seasonal vector form: we can embed seasonal data into an annual vector whose components are the seasonals, and model seasonally heteroscedastic series multivariately. Then fit Vector AR (VAR) models, and compute AR roots from the VAR polynomial's determinant

Contact

Status: R code available for use, paper is halfway complete

Email: tucker.s.mcelroy@census.gov