Seasonal Unit Root Tests in Long Periodicity Cases

David A. Dickey
Department of Statistics, North Carolina State University

1. Introduction

Monthly retail sales and new car sales, among other things, often demonstrate seasonality, that is, a pattern that repeats to some degree on a regular basis with monthly reporting being typical of the two cases just mentioned. Often new car sales are reported along with differences $Y_t–Y_{t–12}$ from the same month of the previous year. Seasonal data may also show an upward trend over time in a healthy growing economy. There are two common kinds of basic model that can be applied to such seasonal data. One approach is a model consisting of a linear time trend and seasonal indicator variables (seasonal dummy variables). This model then assumes that the seasonal effect is the same for all years and equally weights, for example, every January in the computation of the January effect. Another simple model is $Y_t = \alpha Y_{t–12} + e_t$ with e_t being white noise (an independent, constant variance sequence). With this approach, the January effect is predicted each year from its predecessor and every January before that one is ignored. When $\alpha=1$ we have a seasonal random walk in which case it is the differences $Y_t–Y_{t–12}$ rather than the levels that are stationary.

Tables of critical values given for seasonal lags s=2,4, and 12 in Dickey, Hasza, and Fuller (1984, henceforth DHF) can be used with the model $Y_t = \alpha Y_{t–12} + e_t$ to test the hypothesis that $\alpha=1$, that is, the hypothesis that the series is a seasonal random walk and hence must be differenced at a span of 12 to achieve stationarity. When this happens we will say that the series has a seasonal "unit root," a reference to the characteristic polynomial $m^{12}–\alpha$ which has a root 1 when $\alpha=1$. There are approaches, namely Hylleberg, Engle, Granger, and Yoo (1990, often referred to as HEGY) for s=4 and Beaulieau and Miron (1993) for s=12 that consider all 4(12) of the roots separately. We will not address these approaches herein. The DHF article addresses realistic models with additional terms such as linear trend and seasonal dummy variables, however the practical limitation of that article is the small number of s values with available critical values. Other points of note from that article are that the studentized test statistics used for testing have nonnormal distributions even in the limit. As well, the addition of commonly used deterministic terms causes differences even in the limit for the test distributions. Note that reference to the limit here refers to an increasing number of years each having 4 quarters or 12 months. This stands in contrast to what will be done in this article, namely looking at what happens as s is increased. While in reality seasonal periods are fixed, note that for most statistical analyses limit results are used when the sample size is large. In other words, it may be that the asymptotic (s increasing) results will provide good approximations for s values that are not extremely large. Most of the results reported here are worked out in detail in Dickey and Zhang (2010).

2. The Basic Model.

Our model will involve deviations $y_t = Y_t – f(t)$ from some deterministic trend f(t) that would typically contain a seasonal component, like trigonometric functions. The model we have in mind is $y_t = \alpha y_{t–s} + e_t$ where the e's are independent $N(0, \sigma^2)$ error terms, that is, white noise. If $\alpha=1$ then any periodic component of f(t) will cancel out. Further, if f(t) consists of periodic functions and polynomials, the same predictor variables can be use to produce y_t and $y_{t–s}$ from Y_t and $Y_{t–s}$ using regression. Our analysis will assume two steps: first regress Y_t and $Y_{t–s}$ on the predictors then regress residuals on residuals. Regressing Y_t on these predictors and $Y_{t–s}$ gives exactly the same estimate of α and the same test statistics.

We can write time t as t=(j–1)s+i where t is the i^{th} month of year j and s=12 for monthly data. For convenience, we use the terms "year" and "season" here but we can do this for any combination like m days of s=24 hours etc. The variables Y_{ij} and Y_{ik} are seen to be independent for different months j and k in the simple lag s model since, for example, January values depend on only lagged January values and errors as can easily be seen in the upcoming Table 1. Let the e_{ij} and Y_{ij} denote the error and observation at t=(j–1)s+i. If f(t) is periodic of period s then when $\alpha=1$, $Y_t–Y_{t–s} = e_t$ and our model can be conveniently written as $Y_t–Y_{t–s} = (\alpha–1)Y_{t–s} + e_t$ when no deterministic regressors are present, and as

$y_t - y_{t-s} = (\alpha-1)y_{t-s} + e_t$ when deterministic regressors are used. Let the vector Δ be the vector of dependent variables $Y_t - Y_{t-s}$, and let $Y_{(-s)}$ denote the vector of lagged observations. Regress the vector Δ on the matrix X of regressors and a vector $Y_{(-s)}$. The resulting estimator of $\alpha-1$ has numerator $\Delta'[I - X(X'X)^{-1}X']Y_{(-s)}$ and denominator $Y'_{(-s)}[I - X(X'X)^{-1}X']Y_{(-s)}$. Later reference to the "numerator" and "denominator" will refer to these terms, possibly normalized. The terms involving $X(X'X)^{-1}X'$ are "correction terms" for the deterministic predictor variables. These do not appear in the simplest model.

3. The Simplest Estimator

Without any deterministic terms f(t) our estimate of $\alpha-1$, multiplied by $(m(m-1))^{-1/2}$ is the ratio of a numerator term

$$\sum_{i=1}^{s}\left(\sum_{j=1}^{m-1} e_{i,j+1}Y_{ij} \Big/ \left(\sigma^2\sqrt{m(m-1)}\right)\right) = \sum_{i=1}^{s} N_i$$

to a denominator term

$$\sum_{i=1}^{s}\left(\sum_{j=1}^{m-1} Y_{ij}^2\right) \Big/ \left(\sigma^2 m(m-1)\right) = \sum_{i=1}^{s} D_i \ ,$$

where we have used the double subscript notation. Because these are both sums of s independent terms, a central limit theorem will apply as s gets large. Note also that the ratio in question is independent of σ^2. The denominator divided by s converges to its expected value by the weak law of large numbers, and the numerator divided by $s^{1/2}$ converges to a normal random variable. We need only the moments to describe the limit distribution. If the denominator term is multiplied by the regression error mean square and the square root of that product replaces the denominator, that is the studentized test statistic (the regression t statistic) for this simple model and similarly for models with deterministic regressors.

In the text of Fuller (1996) it is shown that the bivariate random variable (N_i, D_i) = $(\sigma^{-2}\sum_{j=1}^{m-1} e_{i,j+1}Y_{ij} / \sqrt{m(m-1)}, \sigma^{-2}\sum_{j=1}^{m-1} Y_{ij}^2 / (m(m-1)))$ converges weakly to $(\frac{1}{2}(W^2(1)-1), \int_0^1 W^2(t)dt)$ where W(t) is a Wiener process on [0,1] and thus when s=1 the regression statistic $\sqrt{m(m-1)}(\hat{\alpha}-1) = \sqrt{m(m-1)}\, N_1/D_1$ converges weakly to $\frac{1}{2}(W^2(1)-1) / \int_0^1 W^2(t)dt$, the error mean square, $\hat{\sigma}^2$, converges in probability to σ^2, and the studentized test statistic, symbolized τ to emphasize its nonstandard distribution, converges to $\frac{1}{2}(W^2(1)-1) / \sqrt{\int_0^1 W^2(t)dt}$.

Note that these are limits with respect to m with s held fixed at 1.

Table 1. Data display with doubly subscripted (t=12(j–1)+i) equivalents.

January (i=1)	February (i=2)	March (i=3) (9 more months follow→)
$Y_1 = e_1$ $(Y_{1,1}=e_{1,1})$	$Y_2=e_2$ $(Y_{2,1}=e_{2,1})$	$Y_3=e_3$ $(Y_{3.1}=e_{3,1})$
$Y_{13}=Y_1+e_{13}=e_1+e_{13}$ $(Y_{1,2}= e_{1,2} + e_{1,1})$	$Y_{14}=Y_2 + e_{14} =e_2 + e_{14}$ $(Y_{2,2}=e_{2,2} + e_{2,1})$	$Y_{15}=Y_3 + e_{15} =e_3 + e_{15}$ $(Y_{3,2}=e_{3,2} + e_{3,1})$
$Y_{25} = e_1+e_{13}+e_{25}$ $(Y_{1,3}= e_{1,3} + e_{1,2} + e_{1,1})$	$Y_{26} = e_2+e_{14}+e_{26}$ $(Y_{2,3}= e_{2,3} + e_{2,2} + e_{2,1})$ (m–3 more rows follow below)	$Y_{27} = e_3+e_{15}+e_{27}$ $(Y_{3,3}= e_{3,3} + e_{3,2} + e_{3,1})$

The January through March data for three years of monthly data are described in Table 1. We assume initial values 0. Note the alternative double subscript notation and the independence of the columns. Were we to observe the full 12 columns and read them read left to right as in a book, the data would be encountered in time order. When read down the columns the data show their monthly random walk nature.

The numerator of $\sqrt{m(m-1)s}(\hat{\alpha}-1)$ for general s can be normalized and written as

$\sum_{i=1}^{s}\left\{\sigma^{-2}\sum_{j=1}^{m-1}e_{i,j+1}Y_{i,j} / \sqrt{m(m-1)}\right\} / \sqrt{s} = \sum_{i=1}^{s}N_i / \sqrt{s}$, a normalized sum of N_i terms as just discussed. The independent identically distributed numerator components N_i have mean $E\{N_i\}$=0 and variance $E\{N_i^2\}$ = ½ so the limit of the normalized numerator as the seasonality s increases is N(0,1/2). The normalized denominator of $\sqrt{m(m-1)s}(\hat{\alpha}-1)$ is $\sum_{i=1}^{s}\left\{\sigma^{-2}\sum_{j=1}^{m-1}Y_{ij}^2 / (m(m-1))\right\} / s = \sum_{i=1}^{s}D_i / s$. Each of these s independent denominator components has mean $E\{D_i\} =$ 1/2 and finite higher order moments. The denominator thus converges in probability to 1/2 and the normalized bias satisfies $\sqrt{m(m-1)s}(\hat{\alpha}-1)\xrightarrow{L}N(0,2)$. As s increases, the mean squared error, $\hat{\sigma}^2$, converges to $(m-1)\sigma^2/(m-2)$ so as s and m increase we have

$$\tau = \sum_{i=1}^{s}N_i / \sqrt{\hat{\sigma}^2\sum_{i=1}^{s}D_i} \xrightarrow{L} N(0,1).$$

This suggests that if the seasonality s increases and m is not too small, the t test in the regression of $Y_t - Y_{t-s}$ on Y_{t-s} (no intercept) could be compared to N(0,1) percentiles whereas it is well known that this is not the case when s=1 (the nonseasonal case). Multiplying τ by $((m-1)/(m-2))^{-1/2}$ would allow use of the standard normal, regardless of m, whenever s is large.

Simulation shows a somewhat bell shaped histogram for τ, though the percentile tables of DHF clearly show τ does not converge to a *standard* normal when s is fixed at 2, 4, or 12 and m increases. The bell shape may be close enough to that of a normal to let the normal distribution be used as a reasonably accurate approximation. This motivates an attempt to center and scale the studentized statistic. To compute the mean and variance needed for the standardization, Taylor's series will be used. The usefulness of the moments thus derived is determined by their performance in simulated series.

4. **Adjusting the mean using Taylor's Series.**

The DHF paper gives τ statistic percentiles for s=2, 4, 12. In the monthly case, for example, the spread between the 5^{th} and 95^{th} percentiles is 3.32 for large sample sizes and is close to that for all their reported sample sizes. The spread between these two percentiles in a standard normal distribution is 3.29 which is close enough to the corresponding standard normal spread 3.32 to suggest that no scale adjustment is necessary. Throughout the DHF tables, the 50^{th} percentiles are close enough to the average of the 5^{th} and 95^{th} to suggest the distribution is close to symmetric about the median, which would imply the mean and median are the same.

A plot of these medians against $s^{-1/2}$ shows points lying very close to a line with slope −1/2 and intercept 0, the median and mean of a standard normal. It appears then that the mean of a Taylor Series approximation to the studentized statistic out through order $s^{-1/2}$ will be close to $-0.5s^{-1/2}$, an observation that we will support analytically. The subtraction of that mean from the studentized statistic without any adjustment for the variance should produce a centered statistic with approximately a N(0,1) distribution, thus allowing tests that involve values of s not tabulated in DHF.

The mean and variance of each numerator term N_i and each denominator term D_i have been discussed earlier but it will also be important to note that these numerator and denominator terms have a covariance approximately 1/3 as given in Dickey (1976), namely

$$\text{Cov}\left(\sigma^{-2}\sum_{j=1}^{m-1}e_{i,j+1}Y_{ij} / \sqrt{m(m-1)}, \sigma^{-2}\sum_{j=1}^{m-1}Y_{ij}^2 / (m(m-1))\right) = (m-2) / (3\sqrt{m(m-1)})$$

The studentized statistic with the error mean square replaced by the true error variance is

$$\tau = \frac{\Sigma N_i}{\sqrt{\Sigma D_i}} = \frac{\sqrt{s}\,\bar{N}}{\sqrt{\bar{D}}} = \sqrt{s}\left[\frac{N_0}{\sqrt{D_0}} + O_p(1/\sqrt{s})\right]$$

where $N_0 = 0$ and $D_0 = ½$ are the means of N_i and D_i. The leading term thus drops out and we become interested in the $O_p(1/\sqrt{s})$ term which will be the key to finding the limit distribution.

Note that if seasonal means are removed from the data or equivalently seasonal dummy variables used in the regression, the expectation of the numerator quadratic form will no longer be 0 thus introducing a nonzero leading term of order $s^{1/2}$ in the expression for τ. The following development thus does not pertain to that case.

Apart from the first, all higher order partial derivatives with respect to N are 0. All partial derivatives purely with respect to D evaluated at (N_0,D_0) are multiples of N_0=0 so they are 0. We thus have

$$\tau = \sqrt{s}\ (\bar{N} - N_0)/\sqrt{D_0} - (1/2)^{-3/2}\sqrt{s}\left(\bar{N} - N_0\right)(\bar{D} - D_0)/2 + O_p(s^{-1}) =$$

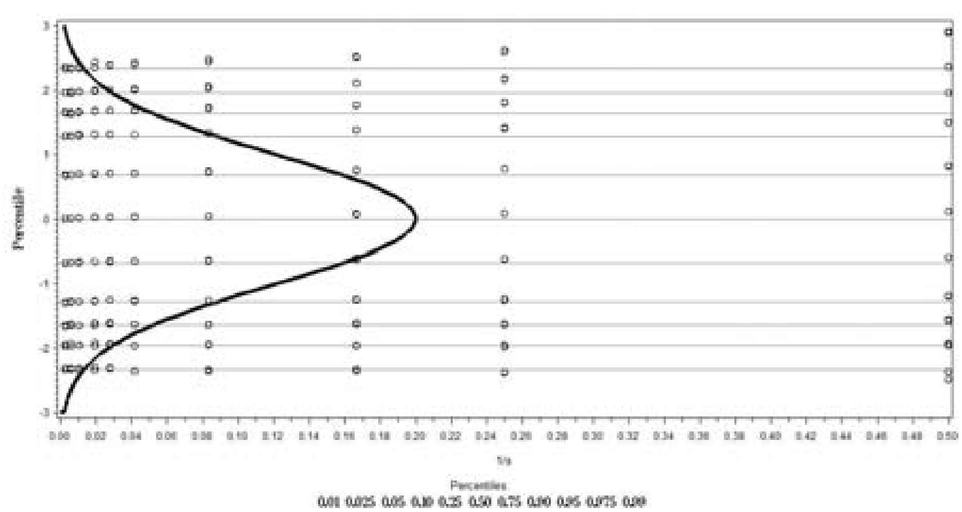
$$\sqrt{2s}\ \bar{N} - (1/2)^{-3/2}\sqrt{s}\,\bar{N}(\bar{D} - 1/2)/2 + O_p(s^{-1})$$

The order 1 term $\sqrt{2s}\ \bar{N}$ is N(0,1) in the limit because $\sqrt{s}\ \bar{N}$ has variance ½. Thus $\tau \xrightarrow{L} N(0,1)$ as s increases, however the DHF tables suggest that s must be rather large for the standard normal distribution to be a good approximation. Further adjustments are needed. The expected value of the term $-(1/2)^{-3/2}\sqrt{s}\,\bar{N}(\bar{D} - 1/2)/2$ is

$$-(2/s)^{1/2}\left(\frac{(m-2)}{3\sqrt{m(m-1)}}\right) \approx -\frac{\sqrt{2}}{3\sqrt{s}} = -0.4714/\sqrt{s} \approx \frac{-1}{2\sqrt{s}}.$$

Centering by subtraction of this from the studentized statistic should bring the expected value to 0 through order $s^{-1/2}$. Figure 1 shows that this approximation is in fact excellent.

Figure 1: Distribution of studentized statistics, adjusted by adding $0.4714/\sqrt{s}$, versus 1/s. The circles mark empirical percentiles typically used in hypothesis testing and the horizontal lines mark the corresponding N(0,1) percentiles. Percentiles are based on two sets of 40,000 simulated series with m=5, making each circle approximately 5 Monte Carlo standard errors in diameter .



5. Adding Periodic Functions of Time.

In the unit root literature it is well known that the addition of deterministic terms, like polynomial trends and/or seasonal means or sinusoids, changes even the limit distribution of the studentized statistic when s=1 or s is any of the values studied in the current literature. When s is allowed to increase some of the problems seen before disappear. The inclusion of seasonal means produces numerator terms N_i that no longer have mean 0 and this case will thus not be dealt with here. If, on the other hand, an overall trend function such as a linear trend is added, simulations in Dickey and Zhang (2010) show that has very little effect when s is large and they show it has no effect in the limit. If a period s function is added (note that the intercept, a column of 1s in regression, repeats at every period and thus is technically periodic) the effect still goes away in large s cases but as with the covariance term earlier, an order $s^{-1/2}$ adjustment helps when s is not extremely large. Interestingly, this adjustment is the same for all period s functions. This section reviews the main ideas behind these statements.

A periodic function f(t) is any function with f(t) = f(t–s) and thus if we regress the seasonal span differences on lagged levels adjusted for the periodic function, the dependent variable will not involve f and for the simplest model $Y_t - f(t) = \alpha(Y_{t-s} - f(t-s)) + e_t$ we will estimate the parameters by first regressing $Y_t - Y_{t-s} = e_t$ on f(t) and Y_{t-s} on f(t) then regressing residuals on residuals which gives exactly the same estimate and test statistic as regressing $Y_t - Y_{t-s}$ on f(t) and Y_{t-s}. We have used the fact that f(t) = f(t–s) and since a collection of periodic predictors, when orthogonalized by the Gramm-Schmidt method, becomes a set of orthogonal periodic functions, we see that if f(t) consists of a linear combination of periodic functions the regressors can be viewed as orthogonal without loss of generality. Looking, then, at a single periodic function for f(t) is sufficient.

A caveat here is that the number k of such deterministic regressors must be fixed or at least satisfy $k/s \rightarrow 0$. For example, if s sines and cosines are used as in Fourier analysis, then the fit is the same as that when s–1 seasonal dummy variables are used and we have noted that the effect of that on the Taylor's series approximation is large enough to invalidate our results here.

Recall from section 2 that the numerator correction term, including normalization, is $\Delta' X (X'X)^{-1} X' Y_{(-s)} / (\sqrt{m(m-1)s}\sigma^2)$. Any periodic regressor, including the intercept, is a column X of n=sm numbers taking on only s values c_i, i=1,2,…,s. The sum of squares for such a column is $X'X = m\sum_{i=1}^{s} c_i^2$.

Assuming $\sum c_i^2 / s$ converges to a positive constant, as for example would be the case with the intercept or sines or cosines, the sum of cross products $\Delta' X$ becomes $\sum_{i=1}^{s} c_i e_{i\bullet} = O_p(\sqrt{sm})$ and, since the lag of Y is a random walk, $X'Y_{(-s)}$ becomes $\sum_{i=1}^{s} c_i Y_{\bullet i} = O_p(\sqrt{sm^3})$. Here the dot notation is used to indicate summation over the associated subscript. The numerator correction term is

$$\Delta' X (X'X)^{-1} X' Y_{(-s)} / \sqrt{m(m-1)s}\sigma^2 = O_p(\sqrt{sm})(m\Sigma c_i^2)^{-1} O_p(\sqrt{sm^3}) / (\sigma^2 \sqrt{m(m-1)s}) = O_p(1/\sqrt{s}) .$$

The correction term can be ignored if s is very large but as happened before, the inclusion of the expectation of this term may add a helpful mean adjustment that allows the normal approximation to work well at less extreme s values.

The normalized denominator has correction term

$$Y'X(X'X)^{-1} X' Y_{(-s)} / (m(m-1)s\sigma^2) = O_p(\sqrt{sm^3})(m\Sigma c_i^2)^{-1} O_p(\sqrt{sm^3}) / (m(m-1)s\sigma^2)$$

which is $O_p(1/s)$ and thus converges to 0 even faster than the numerator correction term as s increases. These facts suggest computing the numerator correction term's expected value then dividing it by the square root of D_0. The expectation in question is

$$E\left\{\left(\sum c_i e_{i\bullet}\right)\left(\sum c_{i'} Y_{i'\bullet}\right)\right\} = \sum c_i^2 E\left\{e_{i\bullet} Y_{i\bullet}\right\}$$

by the independence of the columns in Table 1. Now

$$E\left\{e_{i\bullet} Y_{i\bullet}\right\} = E\left\{\left(\sum_{j=2}^{m} e_{ij}\right)\left(\sum_{j=1}^{m-1}\sum_{k=1}^{j} e_{ik}\right)\right\} = \sigma^2 \frac{(m-1)(m-2)}{2}$$

from which it easily follows that

$$E\left\{\Delta' X (X'X)^{-1} X' Y_{(-s)} / \sqrt{m(m-1)s}\sigma^2\right\} = \left(\frac{(m-1)(m-2)}{2\sqrt{m(m-1)s}} \sum_{i=1}^{s} c_i^2\right)\left(m\sum_{i=1}^{s} c_i^2\right)^{-1} .$$

This does not depend on the values c_i and since the normalized denominator converges to ½ as it did without adjustments, we divide by its square root to get a suggested order $s^{-1/2}$ mean adjustment for the studentized statistic. The suggested adjustment is

$$\left(\frac{(m-1)(m-2)}{2m\sqrt{m(m-1)s}}\right)\sqrt{2} = \left(\frac{\sqrt{m-1}(m-2)}{\sqrt{m^3}}\right)\left(\frac{1}{\sqrt{2s}}\right) \approx \left(\frac{1}{\sqrt{2s}}\right),$$

the last approximation of which holds when m is large. Simulations reported in Dickey and Zhang (2010) show that this adjustment is quite helpful in practice when the number of deterministic regressors k is small relative to s. See that paper for further details, including a proof that the inclusion of any finite number of deterministic regressors has no effect on the limit (s increasing) normal distribution of τ, though as we have seen, s may have to be quite large if further adjustments such as those above are not made.

The following development is taken from Dickey and Zhang (2010) and shows that deterministic polynomial terms added to the model have little effect on the distributions when s and m are fairly large. That paper corroborates the theory with rather large simulations. The basic idea is to separate the year-to-year trend and within year trend. For example a linear trend can be represented as a centered yearly step function with constant increase of the steps plus a within year linear function such that the sum of these two is just the linear function t. By separating the parts, the effect of this variable can be nicely associated with the doubly subscripted variables and the associated sum of squares in the regression correction terms similarly broken down into two components involving variables V and W. The development, taken from Appendix B of that paper, is as follows with some minor modifications to match the syntax of this paper.

For polynomials we will center and scale the entries x_{ij} of the regressor column X obtaining a numerator correction term whose expected value is $E\{(\sum_{i=1}^{s}\sum_{j=1}^{m-1} x_{ij}e_{i,j+1}\sum_{r=1}^{m-1} x_{ir}Y_{ir})/(\sqrt{m(m-1)s}\sum_{i=1}^{s}\sum_{j=1}^{m-1} x_{ij}^2)\}$. We construct X with entries $x_{ij} = W_j + V_i/(m-1)$ and $\sum_{j=1}^{m-1} W_j = \sum_{i=1}^{s} V_i = 0$. For a linear effect, for example, let the centered and scaled (to the interval [–1,1]) variates be $W_j = (2j-m)/(m-1)$ and $V_i = (2i-1-s)/s$ to obtain the linear term

$$x_{ij} = 2\left[\frac{(j-1)s+i}{s(m-1)}\right] - 1 - \frac{1}{s(m-1)} = \left[\frac{2j-m}{m-1}\right] + \left[\frac{2i-1-s}{s}\right]\frac{1}{(m-1)}$$

for j=1,2,…,m–1 and i= 1,2,…,s. Note that $E\{e_{i,j+1}Y_{i,r}\} = \sigma^2$ for j<r and 0 otherwise. The expectation $E\{(\sum_{i=1}^{s}\sum_{j=1}^{m-1} x_{ij}e_{i,j+1}\sum_{r=1}^{m-1} x_{ir}Y_{ir})\}$ is that of a quadratic form $\varepsilon' A\varepsilon = 0.5\varepsilon'(A+A')\varepsilon$ so the expected value is

$$\sigma^2\sum_{i=1}^{s}(\sum_{j=1}^{m-1} x_{ij}\sum_{r=j+1}^{m-1} x_{ir}) = \frac{\sigma^2}{2}\sum_{i=1}^{s}\left(\left[\sum_{j=1}^{m-1} x_{ij}\right]^2 - \sum_{j=1}^{m-1} x_{ij}^2\right) = \frac{\sigma^2}{2}\sum_{i=1}^{s}\left(\left[\sum_{j=1}^{m-1} W_j + V_i\right]^2 - \sum_{j=1}^{m-1} x_{ij}^2\right)$$

and because the sum of the W_j is 0, this is

$$\frac{\sigma^2}{2}\left[\sum_{i=1}^{s} V_i^2 - \sum_{i=1}^{s}\sum_{j=1}^{m-1}(W_j + V_i/(m-1))^2\right] = \frac{\sigma^2}{2}\left[\sum_{i=1}^{s} V_i^2 - s\sum_{j=1}^{m-1} W_j^2 - \sum_{i=1}^{s} V_i^2/(m-1))\right] =$$

$$\frac{\sigma^2}{2}s(m-2)\left[\frac{1}{m-1}\left(\sum_{i=1}^{s} V_i^2/s\right) - \left(\sum_{j=1}^{m-1} W_j^2/(m-2)\right)\right] = O(s) + O(ms).$$

The expectation of the numerator correction term is this quantity divided by $\sqrt{m(m-1)s}\sum_{i=1}^{s}\sum_{j=1}^{m-1} x_{ij}^2$ which grows at rate $m^2s^{3/2}$ so the expectation of the correction term is $O(m^{-1}s^{-1/2})$. With W_j being equally spaced within the

[–1.1] interval, a Rieman sum argument shows that $-\frac{\sigma^2}{4}\left[\sum_{j=1}^{m-1} W_j^2\right]\left(\frac{2}{(m-2)}\right) \to -\frac{\sigma^2}{4}\int_{-1}^{1} x^2 dx = -\frac{\sigma^2}{6}$ and the studentized statistic's mean further involves division by the denominator, thus eliminating the dependence on the error variance. The contribution from the V_i terms is of even lower order. We thus have large sample approximation $-\frac{\sqrt{2}}{6m\sqrt{s}}$ to the mean shift in τ. The coefficient $-\frac{\sqrt{2}}{6m}$ of the $O(s^{-1/2})$ mean shift is decreasing at rate m and quite small even for reasonably small m.

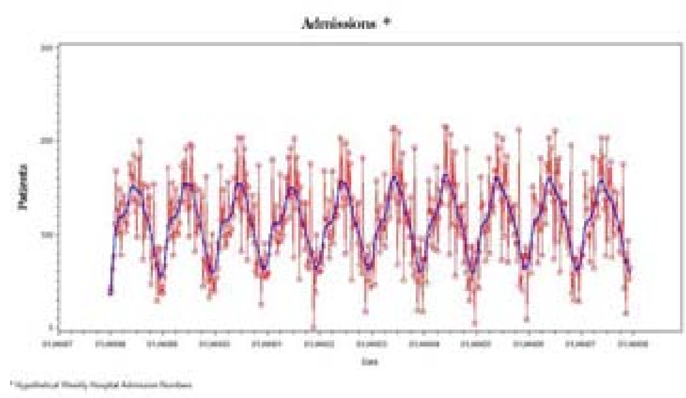
For a quadratic polynomial, we will use $x_{ij} = [(2j-m)/(m-1) + (2i-s-1)/(s(m-1))]^2 - M$ where M is the mean of the squared values appearing before it. This makes the x_{ij} sum to 0. The sum of squares of x_{ij} is increasing at rate md. Now $[(2j-m)/(m-1) + (2i-s-1)/(s(m-1))]^2 = [W_j+V_i/(m-1)]^2$ consists of two pure squared terms, one in W and one in V/(m–1), and a crossproduct which sums to 0 over j. The mean M is therefore the sum of means of the W^2 and $[V/(m-1)]^2$ terms. The effect of centering x_{ij} is the same as that of centering W^2 and $[V/(m-1)]^2$ which then sum to 0 on j and on i respectively. Applying the arguments for the linear trend to this centered quadratic polynomial, the only difference in results is that now the dominant term is $-\frac{\sigma^2}{4}\left[s\sum_{j=1}^{m-1} W_j^4\right]\left(\frac{2}{(m-1)s}\right) \to -\frac{\sigma^2}{4}\int_{-1}^{1} x^4 dx = -\frac{\sigma^2}{10}$ and the approximate mean shift is $-\frac{\sqrt{2}}{10m\sqrt{s}}$, an even smaller shift than in the linear case. It is rare to see polynomials of degree greater than 2 in time series, but the proof extends easily to centered polynomials of any degree. Centering makes this column orthogonal to the intercept which technically is periodic. In practice the added polynomials do not have to be centered or scaled so long as the regression has an intercept. The effect will be the same either way.

6. Example

This study was motivated by a question from a researcher studying hospital admissions for asthma related symptoms. Those data are proprietary so for an example we can look at some generated weekly seasonal data to mimic what might be admissions. Figure 2 is a graph of these hypothetical admissions data with a locally smoothed version overlaid to visualize the seasonal pattern. The seasonality is quite regular so a sine-cosine pair of period 52 is added to the data. The correction needed here for these 3 (including the intercept) deterministic regressors is

$$\frac{\sqrt{2}}{3\sqrt{s}} + \frac{3}{\sqrt{2s}} = \frac{1}{\sqrt{52}}\left(\frac{\sqrt{2}}{3} + \frac{3}{\sqrt{2}}\right) = 0.36$$

Figure 2: Hospital Admissions (hypothetical)



Regressing the differences D on the lag 52 level L, a sine S and cosine C of period 52, we obtain a calculated t statistic t=–1.79 with one-tail P-value 0.0366, obtained by halving the two-tailed P-value from the printed regression output below. With the adjustment we get –1.79 + 0.36 = –1.63>–1.645 which thus has associated one-tail P-value slightly exceeding 0.05. In close calls the adjustment can move the P-value to the other side of the test's level. From the output the estimate of α–1 is –0.02195.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.91778	1.45348	2.01	0.0453
L	1	-0.02195	0.01223	-1.79	0.0733
S	1	0.29552	0.63671	0.46	0.6428
C	1	-0.71018	0.81049	-0.88	0.3814

7. Higher Order Models

A model with $y_t = \alpha y_{t-s} + e_t$ is of somewhat limited use, despite the fact that $y_t = Y_t - f(t)$ which allows for deterministic predictors. Rewriting this model in backshift form with $B(Y_t)=Y_{t-1}$ we have $(1-\alpha B^s)y_t = e_t$ but it may be that that this is not appropriate in that $(1-\alpha B^s)y_t$ is perhaps not white noise but in fact autocorrelated. This can be accommodated by the so-called seasonal multiplicative model $(1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_p B^p)(1-\alpha B^s)y_t = e_t$. DHF give a method for fitting such a model in steps as follows.

Step 1: Fit the model under the null hypothesis, that is, fit an autoregressive order p model to the seasonal differences of y_t where y_t is Y_t adjusted for a regression estimate of f(t). Save the residuals.

Step 2: Estimate the "filtered series" by computing $z_t = (1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_p B^p)y_t$ with the autoregressive coefficients replaced by their estimates. The idea is that for large m this z will approximately follow the simple seasonal lag model already discussed.

Step 3: Regress the residuals from the model in step 1 on lag s of the z_t series and p lags of $y_t - y_{t-s}$ to obtain the studentized test statistic and first order adjustments to the φ_j estimates. The studentized test statistic for z_{t-s} is the same asymptotically (m increasing) as if it were from the simple model. See DHF for details and a formal proof.

8. Conclusion

Tables for testing for seasonal unit roots are available for a few commonly occurring seasonal values such as s=12 for monthly data. For cases where no tables exist and s is large, for example s=52 for weekly data, this paper provides justification for using a normal distribution to calculate approximate p-values. It does so by looking at limits as s increases, finding a rather straightforward central limit theorem. While theoretically pleasing, this result is only practical for very large values of s, however it is neither the variance nor the bell shape of the distribution but rather the mean that is responsible for this drawback. With a mean adjustment, motivated by the Taylor's series expansion of the studentized statistic, that problem is overcome and a statistic that works well in practice even for smaller s values such as 12 and even 4 results. To be practical, the accommodation of a few deterministic terms in the model is needed. Low degree polynomials have little effect on the distribution whereas periodic functions of period s, like sines and cosines, require another simple adjustment of order $s^{-1/2}$ to make the standard normal a reasonable approximation.

References

Beaulieu, J. J. and J.A. Miron, . J.J. Beaulieu and J.A. Miron (1993). Seasonal unit roots in aggregate U.S. data. *Journal of Econometrics* **55**, 305–328.

Dickey, D. A. (1976). Estimation and hypothesis testing in nonstationary time series. Ph.D. thesis Iowa State University.

Dickey, D. A., D. P. Hasza, and W. A. Fuller (1984). "Testing for Unit Roots in Seasonal Time Series", *Journal of the American Statistical Association*, **79**, 355-367.

Dickey, D. A. and Y. Zhang (2010). Seasonal unit root tests in long periodicity cases. *J. Korean Stat. Soc.* **39**, 271-279.

Fuller, W. A. (1996). *Introduction to Statistical Time Series* 2nd ed. Wiley, NY.

Hylleberg S., R.F. Engle, C.W.J. Granger & B.S. Yoo (1990) Seasonal integration and cointegration. *Journal of Econometrics* **44**, 215-238.

Abstract

Dickey, Hasza, and Fuller (1984) gave tables of critical values for seasonal unit root tests. The simplest of these is the studentized statistic (t test) that tests for a coefficient equal to 1 in the regression of the response at time t to the response at time t–s where s is the seasonal lag, s=12 in monthly data for example. As is usual in these so called "unit root tests," the distribution, even in the limit, is non standard, that is, the t statistic is not asymptotically normal. To make matters worse for practical applications, the percentiles change with changes in s and with additional terms, like seasonal means, in the model. Therefore the existing results are limited to the few s values that have been studied, including s=4 and s=12, arguably the most common cases in practice. Herein we take the approach of treating s as an increasing parameter and with this type of asymptotics, find a limit normal distribution, quite in contrast to the fixed s asymptotics previously studied. With some finite sample adjustments, the limit results provide good approximations for surprisingly small s values.

Key Words: Nonstationarity, differencing, asymptotic