|  |
| --- |
| National Bank of Belgium |
| JD+ |
| JD+Seasonality tests |
|  |
| **De Antonio David, Palate Jean** |
| **9/9/2019** |

|  |  |  |
| --- | --- | --- |
| Test | Short description | Implementation classes |
| [1] QS | Test on the seasonal auto-correlations | ec.satoolkit.diagnostics.QsTest,  ec.satoolkit.diagnostics.LjungBoxTest |
| [2] Friedman | Non parametric test (“ANOVA”-type) | ec.satoolkit.diagnostics.FriedmanTest |
| [3] Kruskall-Wallis | Non parametric test on the ranks | ec.satoolkit.diagnostics.KruskallWallisTest |
| [4.1] Auto-regressive  spectrum | Auto-regressive spectrum for identification of seasonal peaks  (Tramo or X12-like) | ec.satoolkit.diagnostics.AutoRegressiveSpectrumTest,  ec.tstoolkit.timeseries.analysis.SpectralDiagnostic |
| [4.2] Tukey spectrum | Tukey spectrum for detection of large seasonal components (Tramo-like) | ec.satoolkit.diagnostics.TukeySpectrumPeaksTest,  ec.tstoolkit.data.BlackmanTukeySpectrum |
| [5] Periodogram | Tests on the sum of a periodogram at seasonal frequencies | ec.satoolkit.diagnostics.PeriodogramTest |
| [6] F-test on seasonal dummies | Estimation of a model with seasonal dummies. Joint F-test on the coefficients of the dummies | ec.satoolkit.diagnostics.FTest |
| “X12” test on seasonality | Combined test on the presence of identifiable seasonality (see Ladiray and Quenneville, 1999) | ec.satoolkit.diagnostics.CombinedSeasonalityTest |
| “TRAMO-SEATS” Seasonality tests | Entry point for several seasonality tests | ec.tstoolkit.modelling.arima.tramo.SeasonalityTests |

Index

1. Scope of this document
2. Seasonality tests in the time domain

AUTOCORRELATION AT SEASONAL LAGS [1]

FRIEDMAN TEST [2]

KRUSKAL-WALLIS TEST [3]

F-TEST ON SEASONAL DUMMIES [6]

1. Seasonality tests in the frequency domain

PERIODOGRAM [5]

AR SPECTRUM [4.1]

TUKEY SPECTRUM [4.2]

1. Annex
2. Lag-window estimators of the spectral density
3. Leakage: Padding and tapering
4. Resolution
5. Equivalence of the tests based on seasonal dummies and on the periodogram in some cases
6. References

Understanding seasonality tests in JD+

1. Scope of this document

The aim of this document is to guide the reader through some of the bibliographical references that contain the concepts and tools required to understand the seasonality tests of JD+, and ultimately, the source code. Some of the tests described ([1], [4], and [6]) are part of the automatic model identification procedure that leads to the seasonal adjustment of the data, but all of them can be used as an independent tool to detect the presence periodical components that are due to both seasonality and calendar effects.

|  |
| --- |
| Figure 1: Seasonality Tests and Analysis |
|  |

The summary results provided in the example of Figure 1 suggest that the series under consideration contains significant seasonal components. Some of those tests rely on alternative estimates of the spectral decomposition of the data. For example, the periodogram, a smoother version of it (Tukey Spectrum) or the autoregressive spectrum, represent complementary approaches to visually detect seasonality and trading day effects. Both are periodic patterns in the data that are highlighted in the graphs using blue and red shadows, respectively. In this document, we describe how this methodology can be translated into formal seasonality tests that can complement the analysis of autocorrelations at seasonal lags.

Finally, it is worth mentioning at this stage that there are groups of tests that can be considered to be equivalent: Friedman [2] and Kruskall-Wallis [3] are practically equivalent by construction, and at the same time, the test on the Periodogram [5] and F-test on seasonal dummies [6], which have different foundations, also turn out to yield very similar results[[1]](#footnote-1). The results of those tests can be complemented with the output of the QS test on seasonal autocorrelations [1] and with both the Autoregressive [4.1] and Tukey [4.2] spectral estimators, which provide a more visual diagnostic. This document also explains that the last two spectrum estimators can complement each other providing analysts with a quick overview of the main seasonal patterns in the data.

1. Seasonality Tests in the Time Domain

AUTOCORRELATION AT SEASONAL LAGS [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

This test checks the correlation between the actual observation and observation lagged by one and two years. In the case of monthly time series the autocorrelation between these values is denoted and respectively. In the case of quarterly time series the autocorrelation between these values is denoted and respectively.

Before describing the QS Statistic proposed by Maravall (2012), let us briefly present a more general formulation, given by the Ljung–Box test. The hypotheses to be tested are:

**HO:** The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process).

**HA:** The data are not independently distributed.

The Ljung–Box test statistic is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where n is the sample size, is the sample autocorrelation at lag , and is the number of lags being tested. Under **HO** the statistic Q follows a distribution. Furthermore, when the test is applied to the residuals of ARIMA(p,q) models, the degrees of freedom need to be adjusted to reflect parameter estimation: . When testing the correlation at lags, we focus on a group of two seasonal correlations. Thus, our Ljung-Box statistic for the case of monthly data would be become:

|  |  |  |
| --- | --- | --- |
|  |  | [1] |

This statistics follows a distribution, where. The p-values are given by. Thus, and would suggest rejecting the null hypothesis at 95% and 99% significance levels, respectively.

Maravall (2012) suggests that a significant Q Statistic can be used to conclude that there is seasonality only when the sign of the autocorrelation coefficients is consistent with such a hypothesis. Thus, negative values of and are replaced by zero, and for. This refinement implies that the presence of seasonality will be detected if any of these conditions holds[[2]](#footnote-2):

* Statistically significant positive autocorrelation at lag 12
* Nonnegative sample autocorrelation at lag 12 and statistically significant

positive autocorrelation at lag 24.

The test’s outcome in our example is displayed as follows:

Figure 2: Test [1]

|  |  |
| --- | --- |
| Tests on autocorrelations at seasonal lags Seasonality present ac(12)=0,6214 ac(24)=0,4508 Distribution: Chi2 with 2 degrees of freedom Value: 93,0661 PValue: 0,000 | Location:  Code: ec.satoolkit.diagnostics.QsTest, ec.tstoolkit.stats.LjungBoxTest  General description:  http://en.wikipedia.org/wiki/Ljung%E2%80%93Box\_test |

FRIEDMAN TEST [2] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The Friedman test is a non-parametric method that will be used to test the significance of the month (or quarter) effect. The test uses the rankings of the observations and does not require distributional assumptions. If the null hypothesis of no stable seasonality is rejected at the 1% significance level then the series is considered to be seasonal and the test’s outcome is displayed in green.

Figure 3: Test [2]

|  |  |
| --- | --- |
| Non parametric (Friedman) test *Based on the rank of the observations in each year* Seasonality present Distribution: Chi2 with 11 degrees of freedom Value: 78,3427 PValue: 0,000 | Location:  Code: ec.satoolkit.diagnostics.FriedmanTest  General description:  <http://en.wikipedia.org/wiki/Friedman_test> |

The test statistic is constructed as follows. Consider first the matrix of data with rows (the blocks, i.e. number of years in the sample), columns (the treatments, i.e. either 12 months or 4 quarters, depending on the frequency of the data). The data matrix needs to be replaced by a new matrix, where the entry is the rank of within block. In other words, is the rank of the period j in the year i.

The test statistic is given by

|  |  |  |
| --- | --- | --- |
|  |  | [2] |

where and. It represents the variance of the average ranking across treatments relative to the total. Under the hypothesis of no seasonality, all months can be equally treated. For the sake of completeness:

* + is the average ranks of each treatment (month) within each block (year)
  + is by construction the average rank

For large n or k, i.e. n > 15 or k > 4, the probability distribution of Q can be approximated by that of a chi-squared distribution. Thus, the p-value is given by.

## Related tests:

* When using this kind of design for a binary response, one instead uses the [Cochran's Q test](http://en.wikipedia.org/wiki/Cochran%27s_Q_test).
* [Kendall's W](http://en.wikipedia.org/wiki/Kendall%27s_W) is a normalization of the Friedman statistic between 0 and 1.
* The [Wilcoxon signed-rank test](http://en.wikipedia.org/wiki/Wilcoxon_signed-rank_test) is a nonparametric test of non-independent data from only two groups.

KRUSKAL-WALLIS TEST [3] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The **Kruskal-Wallis** test is a non-parametric test used for testing whether samples originate from the same distribution. The parametric equivalent of the Kruskal-Wallis test is the [one-way analysis of variance](http://en.wikipedia.org/wiki/One_way_anova) (ANOVA). When rejecting the null hypothesis of the Kruskal-Wallis test, then at least one sample [stochastically dominates](http://en.wikipedia.org/wiki/Stochastic_dominance) at least one other sample. The test does not identify where this stochastic dominance occurs or for how many pairs of groups stochastic dominance obtains. The null hypothesis states that all months (or quarters, respectively) have the same mean. Under this hypothesis the test statistic follows a distribution. When this hypothesis is rejected, it is assumed that time series values differ significantly between periods and the test results are displayed in green:

Figure 4: Test [3]

|  |  |
| --- | --- |
| Non parametric (Kruskal-Wallis) test *Based on the rank of the observations*  Seasonality present Distribution: Chi2 with 11 degrees of freedom Value: 87,3269 PValue: 0,000 | Location:  Code: ec.satoolkit.diagnostics.KruskallWallisTest  General description:  <https://en.wikipedia.org/wiki/Kruskal%E2%80%93Wallis_one-way_analysis_of_variance> |

The test is typically applied to 𝑘 groups of dataeach one is composed of observations which are indexed by.. As before, each month (or quarter) groups all the observations available for a certain number of years. As opposed to the notation used in the Friedman test, number of observations here is not necessarily equal for each group. The ranking of each data point, represented by variable , is now defined different than in Friedman test, since it considers all observables, thereby ignoring group membership. The test statistic is given by, where the numerator and denominator are defined in a slightly different manner than in the Friedman statistic. Here, we have that and, where

* + is the number of observations in group (corresponding to the periods)
  + is the average of the absolute ranks of the data in group
  + is by definition the average rank

Under the null hypothesis that all groups are generated from the same distribution, the test statistic is approximated by a chi-squared distribution. Thus, the p-value is given by. This approximation can be misleading if some of the groups are very small (i.e. less than five elements). If the statistic is not significant, then there is no evidence of stochastic dominance between the samples. However, if the test is significant then at least one sample stochastically dominates another sample.

F-TEST ON SEASONAL DUMMIES [6] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Finally, the test on regression with seasonal dummies checks the presence of deterministic seasonality. The model used here uses seasonal dummies (12 for monthly data, 4 for quarterly data) to describe the logged transformed time series behaviour. The test statistics checks if the seasonal dummies are jointly statistically insignificant. When this hypothesis is rejected, it is assumed that the deterministic seasonality is present and the test results are displayed in green:

Figure 5: Test [6]

|  |  |
| --- | --- |
| Tests on regression with fixed seasonal dummies *Regression model (on original series) with*  *(0 1 1)(0 0 0) noises + mean*  Seasonality present Distribution: F with 11 degrees of freedom in the nominator and 124 degrees of freedom in the denominator Value: 26,6022 PValue: 0,0000 | Location:  Code: ec.satoolkit.diagnostics.FTest |

The GLS-statistic of Lytras, Feldpausch and Bell (2007) requires defining stable or fixed seasonal regressors, . For each one of the 11 regressors, we will have:

|  |  |  |
| --- | --- | --- |
|  |  | [3] |

Those regressors would enter in a regARIMA model, resulting on the following general expression:

|  |  |  |
| --- | --- | --- |
|  |  | [4] |

We always use a (0 1 1)(0 0 0) model in our F-test. The consequences of a misspecified model are discussed in Lytras et al. (2007)

One can use the individual t-statistics to assess whether seasonality for a given month is significant, or a chi-squared test statistic if the null hypothesis is that the parameters are collectively all zero. The chi-squared test statistic is in this case compared to critical values from a -distribution, with degrees of freedom. Since the computed using the estimated variance of may be very different from the actual variance in small samples, this test is corrected using the proposed statistic:

|  |  |  |
| --- | --- | --- |
|  |  | [5] |

where is the sample size, is the degree of differencing and is the total number of regressors in the regARIMA model (including the 11 seasonal dummies and the intercept). This statistic follows a distribution under the null.

1. Seasonality Tests in the Frequency Domain

The identification of **seasonal peaks** can also be carried out with the auto-regressive spectrum and Tukey periodogram. The autoregressive spectrum is based on the estimation of an AR(30) process, while the Tukey periodogram is a non-parametric estimator that introduces some degree of smoothing in the autocovariance generating function. In order to decide whether a series has a seasonal component that is predictable (stable) enough the tests use visual criteria and formal tests that rely on two basic principles: a) the peaks associated to seasonal frequencies should be larger than the median spectrum for all frequencies and, b) the peaks should exceed the spectrum of the two adjacent values by more than a critical value. When such case is detected, the test results are displayed in green. The statistical significance of each one of the seasonal peaks (i.e. frequencies corresponding to 1, 2, 3, 4 and 5 cycles per year) is also displayed.

Figure 6: Test [4]

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Identification of seasonal peaks in a Tukey periodogram and in an auto-regressive spectrum  Seasonality present *T or t for Tukey periodogram, A or a for auto-regressive spectrum; 'T' or 'A' for very signficant peaks, 't' or 'a' for signficant peaks, '-' otherwise*   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **1c/y** | **2c/y** | **3c/y** | **4c/y** | **5c/y** | **6c/y** | | at | At | AT | -t | -- | AT | | Location:  ec.satoolkit.diagnostics.AutoRegressiveSpectrumTest  ec.tstoolkit.timeseries.analysis.SpectralDiagnostic  ec.satoolkit.diagnostics.TukeySpectrumPeaksTest  ec.tstoolkit.data.BlackmanTukeySpectrum |

The so-called **calendar** (trading-day of the week or working days) effects, related to the variance in the number of week-ends per period, can also induce periodic patterns in the data that can be similar to those resulting from pure seasonal effects. In theory, trading day variability is mainly due to the fact that the average number of days in the months or quarters is not equal to a multiple of 7. The average number of days of a month in year of 365.25 days is equal to 365.25/12=30.4375 days, which is unfortunately not a multiple of 7. This means this effect occurs 30.4375/7=4.3482 times per month: one time for each one of the four complete weeks of each month, and a residual of 0.3482 cycles per month, i.e., . This turns out to be a *fundamental frequency* for calendar effects associated to monthly data.

In practice, however, the fact that all months or quarters do not have the same number of working days and trading days generates variability at frequencies that are hard to determine analytically. Other trading day frequencies could also be present in the data, although they tend to be more difficult to detect. Cleveland and Devlin (1980) suggest 0.432 cycles per month for certain flow series and McNulty and Huffman (1989) propose 0.304 cycles per month in order to capture certain patterns of daily activity.

Ladiray (2012) provides an overview of this phenomenon, including the trading day frequencies that are more likely to have an effect on monthly and quarterly data. For the sake of simplicity, Figure 7 below plots the periodogram of two time series that contain the number of working days per month and per quarter, respectively. The peaks, which correspond to calendar effects, provide an intuition regarding the potential role of this phenomenon in actual time series data.

|  |
| --- |
| Figure 7: periodogram of the number of working days (100 years data since 1980) |
|  |
| Note: Blue and red shades highlight frequencies related to seasonal and calendar effects, respectively. The two graphs have been produced with Jdemetra+ (Utilities>>Calendars>>Default). Trading day effects corresponding to each day of the week can also be analyzed with this tool. |

Jdemetra+ introduces the concept of *fundamental* trading day periodicity for all frequencies. As shown in Figure 8, it uses the same logic that has already been described for the case of monthly data, but it also incorporates empirically relevant periodicities for quarterly data, consistent with the pattern resulting from Figure 7.

Figure 8: Defining Trading Day frequencies

|  |
| --- |
| public static double[] getTradingDaysFrequencies(int freq) {  double n = 365.25 // freq is 12 for monthly data,  4 for quarterly data, etc…  double f = 2 \* Math.PI / 7 \* (n - 7 \* Math.floor(n / 7));  if (f > Math.PI) {  f = 2 \* Math.PI - f;  }  if (freq == 12) {  return new double[]{f};  }  else if (freq == 4) {  return new double[]{f, 1.292, 1.850, 2.128};  }  else {  return new double[]{f};  }  } |

This small note regarding calendar effects can be important in our analysis of

seasonality. In particular, the calendar effect in monthly data is very close to the frequency corresponding to 4 cycles per year (red shade in the left panel of Figure 7). This implies that it may be hard to disentangle both effects using the frequency domain techniques discussed below.

PERIODOGRAM [5] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Basic concepts required to understand the tests and to interpret the graphs**

This exposition has been taken from Brockwell and David (1991), but other references such as Bloomfield (2000) or Chatfield (2004) are also suitable.

The vectors constitute an orthonormal basis for

,

The symbol denotes the entire part of

The *Fourier frequencies* are given by multiples of the fundamental frequency

, -<

**Harmonic decomposition**. Any vector can be written as a linear combination of the basis:

,

where is given by all the elements, indexed by *t*, of the vector**:**

Because our harmonic decomposition writes each one of the observations as a linear combination of the elements of the orthonormal basis defined above, the coefficients are unique. Thus, each element of vector can be expressed as a linear combination of harmonics:

*The discrete Fourier transform* of is the sequence of defined above by

*The Periodogram I()* is defined as the squared of the of Fourier transform:

|  |  |  |
| --- | --- | --- |
|  |  | [ 6 ] |

The periodogram can actually be represented using data in by obtaining an orthonormal basis in Rewriting in its **polar form**, where  **is the modulus of**, we can rewrite using an orthonormal basis , as follows:

|  |  |  |
| --- | --- | --- |
|  |  | [ 7 ] |

The basis in that we use to project the data is defined as follows:

* is a vector composed of n elements equal to , which implies that .
* .

This allows us to decompose the sum of squares into components corresponding to and, which are lumped together to produce the “frequency” component for 1 ≥≥.

The coefficients have the form of a simple OLS projection of the data on the orthonormal basis:

|  |  |  |
| --- | --- | --- |
|  |  | [ 8 ] |

Table 1

|  |  |  |
| --- | --- | --- |
| Source | Degrees of freedom | Sum of squares decomposition |
| Frequency | 1 |  |
| Frequency | 2 |  |
|  |  |  |
| Frequency | 2 |  |
|  |  |  |
| Frequency | 1 |  |
| Total | n |  |

The periodogram is then given by the contribution of the harmonic to the total sum of squares, as illustrated by Brockwell and Davis (1991) using the table below:

|  |  |
| --- | --- |
|  | [ 9 ] |

Since are generated by an orthonormal basis, the sum of squares is equal to the sum of the squared coefficients:

|  |  |  |
| --- | --- | --- |
|  |  | [ 10 ] |

Thus:

The periodogram can also be written in terms of the sample autocovariance function for any non-zero Fourier frequency,

|  |  |  |
| --- | --- | --- |
|  |  | [ 11 ] |

and for the zero frequency .

Compare this with the expression for the spectral density of a stationary process:

|  |  |  |
| --- | --- | --- |
|  |  | [ 12 ] |

Chatfield (2004) underlines that is a function of the frequency and not the period, so it would make more sense to use the word “spectrogram” in reference to such function. Some of the expressions and definitions above can be slightly different from those used by other authors. For example, Chatfield (2004) derives the expressions shown above using a slightly different parameterization and proposes to plot the periodogram defined above dividing it by so that the *area* corresponding to the [n/2] intervals defined by the Fourier frequencies, i.e., for all and if is even, can be interpreted as the sum of squares. In JDemetra+, the periodogram object corresponds exactly to the contribution to the sum of squares of the standardized data, since the series are divided by their standard deviation for computational reasons. In turn, Jenkins and Watts (1968) define the periodogram in terms of the variable, which is the approach followed by X-13A-S.

**Defining a -test**

The periodogram defined above for converges in distribution to an exponential function with mean, as shown for example by Theorem 10.3.2 of Brockwell and Davis (1991). It can be shown that is an exponential distribution with mean equal to 2, which is equivalent to a Chi-squared distribution[[3]](#footnote-3) with 2 degrees of freedom,. This implies that a confidence interval for the unobserved spectral density can be given by:

|  |  |  |
| --- | --- | --- |
|  | < , for | [ 13 ] |

A model that is able to represent all features of the data will have residuals that behave like a white noise, the spectral density of which is for all frequencies. Under such null hypothesis we have that, which has been presented in Table 1 as the contribution of the sum of squares associated to frequency, follows a distribution. Thus, we can run the test at seasonal frequencies and reject the null when, is larger than.

If the purpose is to test the joint statistical significance of all seasonal frequencies at the same time, one can consider the closest Fourier frequencies. The test statistic will be given by

**Defining a -test**

Brockwell and Davis (1991, section 10.2) exploit the fact that the periodogram can be expressed as the projection on the orthonormal basis defined in expression [ 8 ] to derive a test including , although the proposed idea could be generalized to non-Fourier frequencies too. Thus, under the null hypothesis:

, for Fourier frequencies

Because is independent from the projection error sum of squares, we can define our F-test statistic as follows:

|  |  |
| --- | --- |
| , for Fourier frequencies | [ 14 ] |

|  |  |
| --- | --- |
|  | [ 15 ] |

where

, and

Thus, we reject the null if our F-test statistic computed at a given seasonal frequency (different from) is larger than. If we consider, our test statistic follows a distribution. The implementation of JDemetra+ considers simultaneously the whole set of seasonal frequencies (cycles per year, being the number of periods per year). Thus, the resulting test-statistic is:

|  |  |
| --- | --- |
|  | [ 16 ] |

where if n is even and 0 otherwise, and

The test performs better in small samples when periodogram is evaluated as the exact seasonal frequencies. Instead of modifying the test, as proposed by Brockwell and Davis (1991), JDemetra+ modifies the sample size to ensure the seasonal frequencies belong to the set of Fourier frequencies. This strategy provides a very simple and effective way to eliminate the so-called problem of leakage that will be described in annex B in more general terms.

AR SPECTRUM [4.1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Definition**

The spectral density at frequency of an AR(p) process with innovation variance is expressed as follows:

|  |  |
| --- | --- |
|  | [ 17 ] |

where denotes the AR(k) coefficient, and

Soukup and Findely (1999) suggest the use of, which in practice much larger than the order that would result from the AIC criterion. The minimum number of observations needed to compute the spectrum is set to for monthly data (or for quarterly series. In turn, the maximum number of observations considered for the estimation is 120. This choice offers enough resolution, being able to identify a maximum of 30 peaks in a plot of 61 frequencies: by choosing, we are able to calculate our density estimates at exact seasonal frequencies (1, 2, 3, 4, 5 and 6 cycles per year). Note that x cycles per year can be converted into cycles per month by simply dividing by twelve, x/12, and to radians by applying the transformation 2.

**Dealing with calendar effects**

As in X-12-ARIMA, the traditional trading day frequency corresponding to 0.348 cycles per month is used in place of the closest frequency. Thus, we replace by. The frequencies neighbouring are set to and.

The proximity of this trading day frequency to the frequency corresponding to 4 cycles per yearcan pose identification problems. The contribution of the trading day frequency may be obscured by the leakage resulting from the potential seasonal peak at, and vice-versa. Representing other trading day frequencies discussed in the literature would imply replacing for byand , respectively (see Figure 9). Jdemetra+ allows the user to modify the number of lags of this estimator and to change the number of observations used to determine the AR parameters. These two options can improve the resolution of this estimator.

Figure 9: Trading Day Frequency for Monthly Data



X-12-ARIMA computes the spectrum for the first differenced input series, the first differences of the seasonally adjusted series, the irregular component and the residuals of a regARIMA model. Soukup and Findely (1999) provide some simulation evidence that suggests that the false alarm rates are reduced by focusing on these residuals. This helps to better isolate unmodeled trading day effects, but also to identify peaks corresponding to seasonal frequencies that are not well captured by the ARIMA model.

**Defining a test**

The statistical significance of the peaks associated to a given frequency is informally tested using a visual criterion, which has proved to perform well in simulation experiments. Visually significant peaks for a frequency satisfy both conditions:

1. , where can be set equal to for all

The first condition implies that if we divide the range in 52 parts or stars, the height of each pick should be at least 6 stars. This measure, used by X-12-ARIMA turns out to yield acceptable calendar effects detection with low false alarm rates. However, Test [4] provided by Jdemetra+ exploit the critical values for (code “A”) and (code “a”) computed by Maravall (2012). He simulates 10000 random walk processes and compute the AR(30) spectrum. All values below the median are set to zero and the resulting for each frequency is ordered. The critical value for a given significance level is simply the first value of that is larger than percent of the 10000 simulated. This implies in practice that percent of the times, we will reject the absence of seasonality or calendar effects (null hypothesis), even when those patterns are present in the data. Maravall (2014) writes that is set for all equal to the critical value associated to the trading frequency,i.e., .

TUKEY SPECTRUM [4.2] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Tukey spectrum is a lag-window estimator**

The Tukey spectrum belongs to the class of lag-window estimators. A lag window estimator of the spectral density

|  |  |
| --- | --- |
|  | [ 18 ] |

is defined as follows:

|  |  |
| --- | --- |
|  | [ 19 ] |

where is the sample autocovariance function, is the ***lag window***, andis the **truncation lag**. is always less than or equal to one, and for. The simple idea behind this formula is to down-weight the autocovariance function for high lags where is more unreliable. This estimator requires choosing as a function of the sample size such that and when . These conditions guarantee that the estimator converges to the true density.

JDemetra+ implements the so-called Blackman-Tukey (or Tukey-Hanning) estimator, which is given by

The choice of large truncation lags decreases the bias, of course, but it also increases the variance of the spectral estimate and decreases the bandwidth (width of the corresponding spectral window. of Brockwell and Davis, 1991). The corresponding spectral window estimator is given by , and corresponds to the Dirichlet kernel:

**Link with spectral window estimators**

Spectral window estimators The Tukey estimator is actually related to the discrete spectral average estimator by defining the ***spectral window***:

Then, the lag window estimator can be written (see page 358 of Brockwell and Davis, 1991) in terms of as follows:

Thus, this expression has the **same form as the discrete spectral average estimator**. Theorem 10.4.1 of Brockwell and Davis (1991) suggest that as

where

|  |  |
| --- | --- |
|  | [ 20 ] |

**Defining a test**

Chatfield (2004) cites Jenkins and Watts (1968, section 6.4.2) for the relationship between truncation lag and the variance of the spectral estimates. This can be justified by theorem 10.3.2 of Brockwell and Davis (1991), regarding the asymptotic properties of the periodogram, which suggests that

Since the lag window estimator can be expressed as a discrete spectral average estimator (definition 10.4.1), the distribution of can be approximated by the distribution of the corresponding linear combination of independent and identically distributed random variables. Tukey (1949) proposes that such distribution can in turn be approximated by with, where and can be estimated by matching the mean and variance of to the asymptotic mean and variance of :

This implies that, where . When and is large, one can also use the approximation shown before:

Because is approximately distributed as, where one can obtain an asymptotic confidence interval for the true density:

|  |  |
| --- | --- |
|  | [ 21 ] |

for . A more general discussion can be found in Chapter 10.5 of Brokwell and Davis (1991), also including a normal approximation.

The calculations for the equivalent degrees of freedom provided in this table are valid for all the lag-window estimators that can be computed with JDemetra+. One could use critical values resulting from the distribution to perform a seasonality test. As summarized by Table 2 the suitable degrees of freedom can be obtained as a function of the lag-window type, the truncation lag, and the sample size (see annex A for an overview of the different types of window).

Table 2

RECTANGULAR OR TRUNCATED WINDOW

TRIANGULAR OR BARTLETT WINDOW

DANIELL WINDOW

BLACKMAN-TUKEY WINDOW (

PARZEN WINDOW

The current JDemetra+ implementation of the seasonality test is not based on the distribution, but on the approximation that has been proposed by Maravall (2012) for TRAMO-SEATS. This test is has been designed for a Blackman-Tukey window based on a particular choices of the truncation lag and sample size. Following this approach, we determine visually significant peaks for a frequency when

,

where is the critical value of a distribution, where the degrees of freedom are determined using simulations. As in TRAMO-SEATS, two significant levels for this test are considered: (code “t” in Test 4) and (code “T” in Test 4). Note that for. As opposed to the AR spectrum, which is computed on the basis of the last 120 data points, we will use here all available observations. Those critical values have been calculated given the recommended truncation lag for a sample size in the interval and for. The approximation is less accurate for sample sizes larger than 300. For quarterly data, , but there are no recommendations regarding the required sample size. Maravall (2012) writes that the purpose of this choice is to identify seasonal peaks from the closest Fourier harmonics and to achieve the necessary resolution to distinguish seasonal from trading day frequencies (see annex C for an example to understand the concept of resolution). As in the case of the AR spectrum, it is expressed in decibel scale:.

Table 3: Tramo-seats combines both spectrum estimators

|  |  |  |
| --- | --- | --- |
|  | **AR spectrum** | **Maravall-Tukey spectrum** |
| **Principle** | Parametric (based on AR approximation) | Non parametric (Blackman-Tukey lag-window estimator) |
| **Visual criterion** | Identifies local peaks that are significant | Identifies significant spectral estimates |
| **Resolution** | It can be improved by using more than 30 lags  Emphasis on exact seasonal and, by adjusting , also trading day frequencies | It can be improved by increasing  Identification of seasonal peaks and trading day effects from the closest Fourier harmonics (not exact seasonal) |
| **Maximum**  **sample size** |  | Full sample |
| **Minimum**  **sample size** |  |  |
| **Test** | X12-ARIMA uses visual criterion (although TRAMO-SEATS proposes a more formal test) | Test based on approximated F distribution with degrees of freedom estimated empirically |
| **Performance** | Accurate peak detection (well documented detection and false alarm rates, also for the trading day frequency) | Tukey spectrum is better than AR at capturing spectral minima due to non-invertibility |

1. Annex
   * + 1. Lag-window estimators

This part describes the main characteristics of the alternative **lag-window estimators** of the spectral density mentioned in Table 1, which can be exploited by the *spectral analysis* tool of JDemetra+. They can be used by defining their parameters in the *properties* window of the so-called Tukey spectrum. All those estimators are defined as a smooth version of the autocovariance generating function, but they are linked to the class of **discrete spectral window estimators**, which is based on the idea of averaging or smoothing the periodogram:

RECTANGULAR OR TRUNCATED WINDOW

The rectangular window is defined as follows . The variance of the resulting lag window estimator converges to when

This lag window estimator approximates the discrete spectral average estimator with weights given by the Dirichlet kernel .

Figure 10



Because is approximately distributed as, where, one can obtain an asymptotic confidence interval for the true density:.

TRIANGULAR OR BARTLETT WINDOW

The triangular window is defined as follows .

The asymptotic variance of the resulting lag window estimator is equal to, which is therefore three times smaller than the one corresponding to the rectangular window.

This lag window estimator approximates the discrete spectral average estimator with weights given by the Fejer kernel . This kernel is displayed in the figure below for values of that equalize the asymptotic variance of to the one corresponding to Dirichlet kernel displayed above.

Figure 11



Because is approximately distributed as, where, one can obtain an asymptotic confidence interval for the true density:.

DANIELL WINDOW

|  |  |
| --- | --- |
| The Daniell window is defined as follows .  The asymptotic variance of the resulting lag window estimator is equal to, which is therefore smaller than the one corresponding to the rectangular window. |  |

This lag window estimator approximates the discrete spectral average estimator with weights given by a kernel with a rectangular shape:

This means that for large values of, the smoothing of the periodogram would be consistent with a very large (constant) weight to the values of the periodogram at a very small neighbourhood of a given Fourier frequency. Conversely, small values of imply a small weight in a large neighbourhood of each Fourier frequency. This rectangular spectral window is related to the concept of **resolution**. The width of this rectangular *spectral* *window* leading to the same asymptotic variance as a given *lag*-*window* estimator is sometimes called the **resolution or bandwidth of the estimator**.

Because is approximately distributed as, where, one can obtain an asymptotic confidence interval for the true density:.

BLACKMAN-TUKEY WINDOW

The Blackman-Tukey window is defined as follows .

The asymptotic variance of the resulting lag window estimator is equal to, which can be smaller or larger than that of the triangular window.

This lag window estimator approximates the discrete spectral average estimator with weights given by the following kernel:

where corresponds to the Dirichlet kernel:

The values and are often referred to as the Tukey-Ha**mm**ing and Tukey-Ha**nn**ing windows, respectively. The last one is implemented in TRAMO-SEATS.

Because is approximately distributed as, where, one can obtain an asymptotic confidence interval for the true density:.

PARZEN WINDOW

The Parzen window is defined as follows

The asymptotic variance of the resulting lag window estimator is equal to , which can be smaller or larger than that of the triangular window.

This lag window estimator approximates the discrete spectral average estimator with weights given by the following kernel:

Because is approximately distributed as, where, one can obtain an asymptotic confidence interval for the true density:.

References:

* Tukey, J. (1949). The sampling theory of power spectrum estimates., Proceedings Symposium on Applications of Autocorrelation Analysis to Physical Problems, NAVEXOS-P-735, Office of Naval Research, Washington, 47-69
* Brockwell, P.J., and R.A. Davis (1991). Times Series: Theory and Methods. Springer Series in Statistics.
  + - 1. Leakage: Padding and Tapering

The periodogram is an asymptotically unbiased estimator of the spectral density (see for example proposition 10.3.1. of Brockwell and Davis, 1991). Nevertheless, as it will be illustrated below, the periodogram can be misleading in small samples, since it can attribute non-zero explanatory power to periods that do not exist in the data.

Consider for example data generated by the following deterministic process:

Figure 12: Deterministic function with periodicity of 12 months



Note that the periodogram displayed in Figure 13 for sample size equal to , which is a multiple of the annual frequency of the data, is equal to zero for all frequencies except. In turn, when observations are used, the periodogram is different from zero also in the neighbourhood of. This implies that frequencies other than contribute to account for the variance in even if the data generating process is. This problem is known as leakage and it typically discussed in the context of lag-window estimators of the spectral density, since they are by construction weighted averages of the periodogram.

Figure 13: Periodogram of , for and for



Although any period that is a multiple of 12 guarantees that the seasonal frequencies belong to the set of Fourier frequencies, this is not sufficient to include other periodic sources of variation such as trading day effects. However, considering 28 years are in a sample of monthly data implies using the first entire multiple of the fundamental trading day period.

If we are willing to ignore leap-year effects, and imagine that one year is composed of 365 days (instead of 365.25 days), a sample of size equal to the first entire multiple of that theoretical trading day period would be obtained with only 7 years of data:

This problem is known as leakage, and we have already seen in the previous section that the periodogram computed with a sample size that is a multiple of 12 is equal to zero for all Fourier frequencies different from . Thus, the problem of leakage is not visible if we simply restrict the analysis to the Fourier frequencies. However, the use of a finer grid, our perception of reality is blurred, since many other frequencies in the neighbourhood of are now contributing to account for the variance in the data.

Figure 14: Periodogram of , for and for



The strategy of adding observations (typically zeros) in order to make sure the Fourier frequencies incorporate certain target frequencies is known as ***padding***. JDemetra+ follows the same principal, but instead of adding artificial data, we remove data points so that is a multiple of the annual frequency. As a result, the seasonal frequencies are contained in the set of Fourier frequencies.

More generally, spectral average estimators, by definition, imply that large peaks corresponding to a certain frequencies can “leak” into other frequencies. Bloomfield (2000) shows that leakage in the neighbourhood of any frequency can be effectively reduced by applying a split cosine bell function to the data, which is mathematically equivalent to the so-called method of ***hanning***.

This split cosine bell is implemented in JDemetra+ and can be used by opening the Properties window. The so-called ***taper*** parameter refers to the percentage of observations that will be transformed. For values of *p>*0, the ends of the data (p/2 at each side) are gradually set to zero. In the limit, *p=*0, whole time series is left untransformed, i.e. it is simply multiplied by 1, which implies that the periodogram corresponds exactly to the *Fourier transform* of the data. Thus, our finite sample can be interpreted as the outcome of having applied a rectangular window to the whole population. In the other extreme, if *p=*100%, the resulting periodogram can be referred to as the *hanned transform* of the data. Tukey (1967) has suggested p=10% or 20%.

As illustrated in Figure 15, which plots the Tukey spectrum of 138 monthly observations referring to German exports, tapering the whole series (p=1) does does not seem to provide a different picture. Although the spectrum at the trading day frequency of resulting from tapering is slightly higher, both pictures provide a very similar picture. In order to actually improve the resolution so that the trading day frequency of and the seasonal frequency of can be better isolated, it is more effective to change the window length, as explained in the next section.

|  |
| --- |
| Figure 15: Tukey spectrum for German exports |
|  |
|  |
|  |
|  |

* + - 1. Resolution

If two periodic components in the series are close to the same frequency, smoothing the periodogram might be incapable of identifying, or resolving, the individual peaks. The bandwidth of a spectral estimate is related with the frequency interval applicable to smooth the periodogram. However, such a definition of bandwidth would neglect the fact that the periodogram, although asymptotically unbiased, is not a consistent estimator of the spectral density. Here, we define the bandwidth or resolution for any lag-window spectral estimator as the width of a rectangular spectral window

|  |  |  |
| --- | --- | --- |
|  |  | [22 ] |

that achieves the same asymptotic variance.

Consider for example the Tukey spectrum, which has an asymptotic variance equal to. We know that for large values of the truncation lag , the smoothing of the periodogram would be consistent with the relatively larger weight to the values of the periodogram that are closest to a given Fourier frequency, using the weights given by a *bell-shaped* curve. However, in order to define the resolution or bandwidth of this estimator for given values of, we should ask the following question: what is the width a *rectangular* spectral window, i.e., according to expression [22], leading to the same asymptotic variance to? By equation this value to the asymptotic variance resulting from the rectangular spectral window , we obtain , which implies that the bandwidth is . Maravall (2014) proposes to use the Tukey window with , which is the smallest value so that the trading day frequency of and the seasonal frequency of can be resolved. This implies . Although this is larger than the distance between, the identification problem could be solved thanks to the bell-shape of the spectral window, the weights of which are not constant as in the rectangular spectral window. In the example of Figure 16, where the sample size is larger than 120, it turns out that a window length equal is sufficient to define a high value of the spectrum around both frequencies. However, Maravall recommends to use when the sample size is larger than 120, as in our example. This choice implies a higher resolution and it is translated in the identification of two separate peaks for the seasonal and trading day frequencies.

To conclude, the concept of resolution should not be confused with the number of elements used to form the frequency grid, although it is clear that higher resolution windows can benefit from finer grids. This concept can also be used in the framework of the AR spectrum. By increasing the number of lags in the AR model, higher resolution estimates of the spectral can be obtained.

|  |
| --- |
| Figure 16: Tukey spectrum for German exports |
|  |
|  |
|  |

D- Equivalence between periodogram test [5] and seasonal dummies test [6]

Both tests provide identical results when the number of observations is a multiple of its annual frequency, so that the seasonal frequencies coincide with Fourier frequencies and when the F-test is computed on an OLS model with mean (or equivalently, when the estimated (0 1 1)(0 0 0) model degenerates to such a model (ma parameter ≈ -1)). The coefficients associated to seasonality are then obtained from the linear projection of the data on the same sub-space.

Let the data, the periodogram at the seasonal frequencies is given by coefficients associated to the orthonormal basis onto which the data is projected, i.e. 2 coefficients corresponding to each frequency given by given by , plus one coefficient for :

where .

In turn, the F-Test is obtained by projecting the data on the orthonormal space given by the following basis (or a linear combination of it):

where if ( and otherwise.

Both orthonormal bases have the same dimensions, and it turns out that all the elements of one basis can be generated by the other one. It can be shown that

and

by choosing and . The relation holds for . Taking into account that , , and, the result is immediate.

Since we can express the regression on seasonal dummies (+ mean) as a regression on cos/sin at the frequencies corresponding to the seasonal frequencies, the resulting F-tests are identical.

In both cases, the null hypothesis is that the data is a white noise. So, the tests are well suited to the residuals of an ARIMA model.

1. References

* Bloomfield, P. (2000) *Fourier Analysis of Time Series: An Introduction*. Wiley-Interscience
* Brockwell, P.J., and R.A. Davis (1991). Times Series: Theory and Methods. Springer Series in Statistics.
* Chatfield, C. (2004) *The Analysis of Time Series: An Introduction*. Chapman &Hall/CRC
* Cleveland, W.S. and Devlin, S.J., (1980), "Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods," Journal of the American Statistical Association, Vol. 75, No. 371, pp. 487-496.
* [Friedman, Milton](http://en.wikipedia.org/wiki/Milton_Friedman) (December 1937). "The use of ranks to avoid the assumption of normality implicit in the analysis of variance". Journal of the American Statistical Association (American Statistical Association) 32 (200): 675–701. [doi](http://en.wikipedia.org/wiki/Digital_object_identifier):[10.2307/2279372](http://dx.doi.org/10.2307%2F2279372). [JSTOR](http://en.wikipedia.org/wiki/JSTOR) [2279372](http://www.jstor.org/stable/2279372).
* [Friedman, Milton](http://en.wikipedia.org/wiki/Milton_Friedman) (March 1939). "A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance". Journal of the American Statistical Association (American Statistical Association) 34 (205): 109. [doi](http://en.wikipedia.org/wiki/Digital_object_identifier):[10.2307/2279169](http://dx.doi.org/10.2307%2F2279169). [JSTOR](http://en.wikipedia.org/wiki/JSTOR) [2279169](http://www.jstor.org/stable/2279169).
* [Friedman, Milton](http://en.wikipedia.org/wiki/Milton_Friedman) (March 1940). "A comparison of alternative tests of significance for the problem of m rankings". The Annals of Mathematical Statistics 11 (1): 86–92. [doi](http://en.wikipedia.org/wiki/Digital_object_identifier):[10.1214/aoms/1177731944](http://dx.doi.org/10.1214%2Faoms%2F1177731944). [JSTOR](http://en.wikipedia.org/wiki/JSTOR) [2235971](http://www.jstor.org/stable/2235971).
* Ljung G. M. and G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models". [Biometrika](http://en.wikipedia.org/wiki/Biometrika) 65 (2): 297–303. [doi](http://en.wikipedia.org/wiki/Digital_object_identifier):[10.1093/biomet/65.2.297](http://dx.doi.org/10.1093%2Fbiomet%2F65.2.297)
* Lytras, D. P., Feldpausch, R. M., and Bell, W. R. (2007), “Determining Seasonality: A Comparison of Diagnostics from X-12-ARIMA,” Proceedings of the Third International Conference on Establishment Surveys, Montreal, Canada
* Jenkins, G.M. and D.G. Watts (1968) *Spectral Analysis and its Applications*. San Francisco: Holden-Day.
* Kruskal; Wallis (1952). "Use of ranks in one-criterion variance analysis". [Journal of the American Statistical Association](http://en.wikipedia.org/wiki/Journal_of_the_American_Statistical_Association) 47 (260): 583–621. [doi](http://en.wikipedia.org/wiki/Digital_object_identifier):[10.1080/01621459.1952.10483441](http://dx.doi.org/10.1080%2F01621459.1952.10483441).
* Ladiray, D. (2012), “Theoretical and Real Trading Day Frequencies”. Economic Time Series: Modeling and Seasonality, Edited by Tucker S., McElroy Chapman and Hall, pages 255-279, DOI: 10.1201/b11823-16
* Ladiray, D. and B. Quenneville (1999), “Seasonal Adjustment with the X-11 method”. Lecture Notes in Statistics, New York: Springer-Verlag
* Leemis, L.M. and J.T. McQueston (2008) “Univariate Distribution Relationships”. American Statistical Association Vol. 62, No. 1, page 45-53 DOI: 10.1198/000313008X270448
* Maravall (2011).”Seasonality Tests and Automatic Model Identification in Tramo-Seats”. Manuscript (the version we have)

or

* Maravall, A. (2012). “Update of Seasonality Tests and Automatic Model Identification in TRAMO-SEATS”. Bank of Spain (November 2012 draft) (version cited by Findley)
* McNulty, M.; Huffman, W. (1989) [The Sample Spectrum of Time Series with Trading Day Variation](http://www.econ.iastate.edu/research/journal-articles/p4850), Economics Letters Vol. 31 (1989): 367-370.
* Soukup, R.J., and D.F. Findley (1999) On the Spectrum Diagnosis used by X12-ARIMA to Indicate the Presence of Trading Day Effects After Modeling or Adjustment. In *Proceedengs of the American Statistical Association. Business and Economic Statistics Section, 144-149, Alexandria, VA.*
* Tukey, J. (1949). The sampling theory of power spectrum estimates., Proceedings Symposium on Applications of Autocorrelation Analysis to Physical Problems, NAVEXOS-P-735, Office of Naval Research, Washington, 47-69

1. It can be shown that there are strictly equivalent in some special cases. See annex D [↑](#footnote-ref-1)
2. Taking into account the modification of the test, the distribution is no longer a; the critical values should also be adapted. [↑](#footnote-ref-2)
3. See for example Leemis, L.M. and J.T. McQueston (2008) “Univariate Distribution Relationships”. American Statistical Association Vol. 62, No. 1, page 45-53 DOI: 10.1198/000313008X270448 [↑](#footnote-ref-3)