JDemetra+ online documentation

Anna Smyk, Alain Quartier-la-Tente, Tanguy Barthelemy, Karsten Webel

12/23/2022

Table of contents

Preface

Welcome to the JDemetra+ online documentation.

JDemetra+ is a software for seasonal adjustment and other time series functions, developed in the framework of Eurostat's "Centre of Excellence on Statistical Methods and Tools" by the National Bank of Belgium with the support of the Bundesbank.

To learn more about this project https://ec.europa.eu/eurostat/cros/content/centre-excellence-statistical-methods-and-tools.

1 JDemetra+

1.1 Introduction

JD+ is a library of algorithms for seasonal adjustment and time series econometrics. You can learn more about the history of the project here (link to below)

link to key references - handbooks (3) - sets of guidelines (2 or 3 ?)

1.2 Main functions

1.2.1 Seasonal adjustment algorithms

All are available for low and high frequency data.

Algorithms	Access	Key features
X13-Arima		
Tramo-Seats		
STL		
State Space Models (STS)		

1.2.2 Temporal Disaggregation and benchmarking

Algorithms	Access	Key features
Chow-lin		
Fernandez		

1.2.3 Trend-cycle estimation

Algorithms	Access	Key features

1.3 Nowcasting

Algorithms	Access	Key features

1.4 Structure of this book

This book is divided in four parts, allowing the user to access the resources from different perspectives.

1.4.1 Algorithms

This part provides a step by step description of all the algorithms featured in JD+, grouped by purpose - seasonal adjustment - benchmarking - temporal disaggregation - … links

1.4.2 Tools

JDemetra+ offers 3 kinds of tools - A Graphical User Interface (GUI) which can be enhanced with plug-ins - A set of R packages - A Cruncher for mass production in seasonal adjustment

1.4.3 Underlying Statistical Methods

This part gives details about the underlying statistical methods to foster a more in-depth understanding of the algorithms.Those methods are described in the light and spirit of their use as building blocks of the algorithms presented above, not aiming at all at their comprehensive coverage.

1.5 How to use this book

1.5.1 Audience

This book targets the beginner as well as seasoned methodologist interested in using JDemetra+ software for any the purposes listed below.

The documentation is built in layers allowing to skip details and complexity in the first steps

2 Seasonal Adjustment

2.1 Motivation

The primary aim of the seasonal adjustment process is to remove seasonal fluctuations from the time series.

[insert def SA from b_ov]

2.2 Data frequencies

The seasonal adjustment methods available in JDemetra+ aim to decompose a time series into components and remove seasonal fluctuations from the observed time series. The X-11 method considers monthly and quarterly series while SEATS is able to decompose series with 2, 3, 4, 6 and 12 observations per year.

2.3 Unobserved Components (UC)

The main components, each representing the impact of certain types of phenomena on the time series (X_t), are:

- The trend (T_t) that captures long-term and medium-term behaviour;
- The seasonal component (S_t) representing intra-year fluctuations, monthly or quarterly, that are repeated more or less regularly year after year;
- The irregular component (I_t) combining all the other more or less erratic fluctuations not covered by the previous components.

In general, the trend consists of 2 sub-components:

- The long-term evolution of the series;
- The cycle, that represents the smooth, almost periodic movement around the long-term evolution of the series. It reveals a succession of phases of growth and recession.

To achieve this goal, seasonal adjustment methods decompose the original time series into components that capture specific movements. These components are: trend-cycle, seasonality and irregularity. The trend-cycle component includes long-term and medium-term movements in the data. For seasonal adjustment purposes there is no need to divide this component into two parts. JDemetra+ refers to the trend-cycle as trend and consequently this convention is used here. For seasonal adjustment purposes both TRAMO-SEATS and X-13ARIMA-SEATS do not separate the long-term trend from the cycle as these two components are usually too short to perform their reliable estimation. Consequently, hereafter TRAMO-SEATS and X-13ARIMA-SEATS estimate the trend component. However, the original TRAMO-SEATS may separate the long-term trend from the cycle through the Hodrick-Precsott filter using the output of the standard decomposition. It should be remembered that JDemetra+ refers to the trend-cycle as trend (T_t), and consequently this convention is used in this document.

TRAMO-SEATS considers two decomposition models:

- The additive model: $X_t = T_t + S_t + I_t$;
- The log additive model: $log(X_t) = log(T_t) + log(S_t) + log(I_t)$.

Apart from these two decomposition types X-13ARIMA-SEATS allows the user to apply also the multiplicative model: $X_t = T_t \times S_t \times I_t$.

A time series x_t, which is a subject to a decomposition, is assumed to be a realisation of a discrete-time stochastic, covariance-stationary linear process, which is a collection of random variables x_t, where t denotes time. It can be shown that any stochastic, covariance-stationary process can be presented in the form:

$x_t = \mu_t + \tilde{x}_t$,

1

where μ_t is a linearly deterministic component and $\tilde{x}_t$ is a linearly interderministic component, such as:

$$\tilde{x}_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

,

2

where $\sum_{j=0}^{\infty} \psi_i^2 < \infty$ (coefficients ψ_j are absolutely summable), $\psi_0 = 1$ and a_t is the white noise error with zero mean and constant variance V_a. The error term a_t represents the one-period ahead forecast error of x_t, that is:

$$a_t = \tilde{x}_t - \hat{x}_{t|t-1}$$

,

3

where

$$\hat{x}_{t|t-1}$$

is the forecast of

$$\tilde{x}_t$$

made at period $t-1$. As a_t represents what is new in

$$\tilde{x}_t$$

in point t, i.e., not contained in the past values of

$$\tilde{x}_t$$

, it is also called innovation of the process. From

3

$$\tilde{x}_t$$

can be viewed as a linear filter applied to the innovations.

The equation 7.1 is called a Wold representation. It presents a process as a sum of linearly deterministic component μ_t and linearly interderministic component $\sum_{j=0}^{\infty} \psi_j a_{t-j}$, the first one is perfectly predictable once the history of the process x_{t-1} is known and the second one is impossible to predict perfectly. This explains why the stochastic process cannot be perfectly predicted.

Under suitable conditions

$$\tilde{x}_t$$

can be presented as a weighted sum of its past values and a_t, i.e.:

$$\tilde{x}_t = \sum_{j=0}^{\infty} \pi_j \tilde{x}_{t-j} + a_t$$

,

4

In general, for the observed time series, the assumptions concerning the nature of the process

1

do not hold for various reasons. Firstly, most observed time series display a mean that cannot be assumed to be constant due to the presence of a trend and the seasonal movements.

Secondly, the variance of the time series may vary in time. Finally, the observed time series usually contain outliers, calendar effects and regression effects, which are treated as deterministic. Therefore, in practice a prior transformation and an adjustment need to be applied to the time series. The constant variance is usually achieved through taking a logarithmic transformation and the correction for the deterministic effects, while stationarity of the mean is achieved by applying regular and seasonal differencing. These processes, jointly referred to as preadjustment or linearization, can be performed with the TRAMO or RegARIMA models. Besides the linearisation, forecasts and backcasts of stochastic time series are estimated with the ARIMA model, allowing for later application of linear filters at both ends of time series. The estimation performed with these models delivers the stochastic part of the time series, called the linearised series, which is assumed to be an output of a linear stochastic process.[1] The deterministic effects are removed from the time series and used to form the final components.

In the next step the linearised series is decomposed into its components. There is a fundamental difference in how this process is performed in TRAMO-SEATS and X-13ARIMA-SEATS. In TRAMO-SEATS the decomposition is performed by the SEATS procedure, which follows a so called ARIMA model based approach. In principle, it aims to derive the components with statistical models. More information is given in the SEATS section. X-13ARIMA-SEATS offers two algorithms for decomposition: SEATS and X-11. The X-11 algorithm, which is described in the X-11 section section, decomposes a series by means of linear filters. Finally, in both methods the final components are derived by the assignment of the deterministic effects to the stochastic components. Consequently, the role of the ARIMA models is different in each method. TRAMO-SEATS applies the ARIMA models both in the preadjustment step and in the decomposition procedure. On the contrary, when the X-11 algorithm is used for decomposition, X-13ARIMA-SEATS uses the ARIMA model only in the preadjustment step. In summary, the decomposition procedure that results in an estimation of the seasonal component requires prior identification of the deterministic effects and their removal from the time series. This is achieved through the linearisation process performed by the TRAMO and the RegARIMA models, shortly discussed in the Linearisation with the TRAMO and RegARIMA models section.The linearised series is then decomposed into the stochastic components with SEATS or X-11 algorithms.

X-13

X-13ARIMA is a seasonal adjustment program developed and supported by the U.S. Census Bureau. It is based on the U.S. Census Bureau's earlier X-11 program, the X-11-ARIMA program developed at Statistics Canada, the X-12-ARIMA program developed by the U.S. Census Bureau.

[1]When the series are non-stationary differentiation is performed before the seasonality tests.

2.4 Detecting seasonal patterns

2.5 Pre-treatment

2.5.1 Calendar correction

details of regressor building in calendar chapter

2.5.1.1 rationale

2.5.1.2 method

2.5.1.3 tools

2.5.2 Outliers

2.5.2.1 rationale

2.5.2.2 method

2.5.2.3 tools

2.5.3 Reg-Arima Model

Tramo and Reg-Arima are very similar…details in M chapter

2.5.4 Model evaluation

goodness-of-fit

2.6 X-11 Decompostion

this part should allow to use x-11 via RJDemetra as well as via GUI

2.6.1 Quick launch with default specifications

2.6.2 Output 1: series

X-11 gives access to a great part of it's intermediate computations

Here we focus on the final components (Table D)

List of series (edit : table with name and meaning)

Retrieve in GUI

Retrieve in R

2.6.3 Output 2: final parameters

Relevant if parameters not set manually

List Final trend filter final seasonal filer

Retrieve via GUI image

Retrieve in R

2.6.4 Output 3: diagnostics

X11 produces the following type diagnostics or quality measures Table with link to detail

2.6.5 Specifications / parameters

2.6.5.1 List

2.6.5.1.1 General settings

- **Mode**
  - check if this option still works, if so add and edit instructions from old page)
  - if not but button present : explain that the mode is determined in pre-adjustment (function)
- **Seasonal component**
  - check if still relevant, idem as above
  - in v.2.3 if not ticked, S estimated but options on seasonal filter not available
- **Forecasts horizon**

Length of the forecasts generated by the RegARIMA model - in months (positive values) - years (negative values) - if set to is set to 0, the X-11 procedure does not use any model-based forecasts but the original X-11 type forecasts for one year. - default value: -1, thus one year from the Arima model

- **Backcasts horizon**

Length of the backcasts generated by the RegARIMA model - in months (positive values) - years (negative values) - default value: 0

2.6.5.1.2 Irregular correction

- **LSigma**
  - sets lower sigma (standard deviation) limit used to down-weight the extreme irregular values in the internal seasonal adjustment iterations, learn more here (LINK to M_ chapter)
  - values in $[0, Usigma]$
  - default value is 1.5
- **USigma**
  - sets upper sigma (standard deviation)
  - values in $[Lsigma, +\infty]$
  - default value is 2.5
- **Calendarsigma**
  - allows to set different **LSigma** and **USigma** for each period
  - values
    - None (default)
    - All: standard errors used for the extreme values detection and adjustment computed separately for each calendar month/quarter
    - Signif: groups determined by cochran test (check)
    - Sigmavec: set two customized groups of periods
- **Excludeforecasts**
  - ticked : forecasts and backcasts from the RegARIMA model not used in Irregular Correction
  - unticked (default): forecasts and backcasts used

2.6.5.1.3 Seasonality extraction filters

- **Seasonal filter** choice

Specifies which be used to estimate the seasonal factors for the entire series (link to relevant part in M chapter)

- S3 $\times$ 1 – 3 $\times$ 1 moving average.
- S3 $\times$ 3 – 3 $\times$ 3 moving average.
- S3 $\times$ 5 – 3 $\times$ 5 moving average.
- S3 $\times$ 9 – 3 $\times$ 9 moving average.
- S3 $\times$ 15 – 3 $\times$ 15 moving average.
- Stable – a single seasonal factor for each calendar period is generated by calculating a simple average over all values for each period (taken after detrending and outlier correction).
- X11Default – 3 $\times$ 3 moving average is used to calculate the initial seasonal factors and a 3 $\times$ 5 moving average to calculate the final seasonal factors.
- Msr – automatic choice of a seasonal filter. The seasonal filters can be selected for the entire series, or for a particular month or quarter.
- default value: Msr

Check: will user choice be applied to all steps or only to final phase D step

- **Details on seasonal filters**

Sets different seasonal filters by period in order to account for seasonal heteroskedasticity (link to M chapter)

- default value: empty

2.6.5.1.4 Trend estimation filters

- **Automatic Henderson filter** our user-defined
  - default: length 13
  - unticked: user defined length choice
- **Henderson filter** length choice
  - values: odd number in $[3, 101]$
  - default value: 13

Check: will user choice be applied to all steps or only to final phase D step

2.6.5.2 Parameter setting in GUI

here v2, adjust to v3 asap

2.6.5.3 Parameter setting in R packages

extensive help on functions available in package help pages Rcode snippets

In R, to implement any param change, it is required to retrieve current spec, modify it and apply it again (see T R packages chapter for details). (specific link)

here example changing all the settings (just remove irrelevant changes)

Rjdemetra (v2) Edit : here static R code link to a "worked example" wih dynamic code ticked box in GUI corresponds to …? in R

```
#Creating a modified specification, customizing all available X11 parameters
modified_spec<- x13_spec(current_sa_model,
    #x11.mode="?",
    #x11.seasonalComp = "?",
    x11.fcasts = -2,
    x11.bcasts = -1,
    x11.lsigma = 1.2,
    x11.usigma = 2.8,
    x11.calendarSigma = NA, # EDIT with example
      x11.sigmaVector = NA,
    x11.excludeFcasts = NA
    # filters
    x11.trendAuto = NA, # needed inf value ?
    x11.trendma = 23,
    x11.seasonalma = "S3X9
    # details on seasonal filters)

#New SA estimation : apply modified_spec
modified_sa_model<-x13(raw_series,modified_spec)
```

EDIT : link to package help page v2 +v3

2.7 STL

2.8 SEATS

2.9 SSF

2.10 Quality assessment

2.10.1 Residual seasonality

2.10.2 Residual calendar effects

3 Seasonal adjustment of high frequency data

3.1 Motivation

3.1.1 Ubiquitous use

3.1.2 Data specificities

3.2 Tools

code here and/or link to R packages chapter

3.3 Unobserved Components

3.4 Identifying seasonal patterns

3.4.1 Spectral analysis

3.4.2 Seasonality tests

3.5 Pre-adjustment

3.5.1 Calendar correction

3.5.2 Outliers and intervention variables

3.5.3 Linearization

3.6 Decomposition

3.6.1 Extended X-11

3.6.2 STL decomposition

3.6.3 Arima Model Based (AMB) Decomposition

3.6.4 State Space framework

3.7 Quality assessment

3.7.1 Residual seasonality

3.7.2 Residual Calendar effects

4 Outlier detection

(in or outside a seasonal adjustment process)

4.1 Motivation

4.2 With Reg Arima models

4.3 Specific TERROR tool

4.4 With structural models (BSM)

5 Calendar and user-defined corrections

This chapter describes the generating process of calendar regressors, outliers and other input variables. The use of this variables inside a seasonal adjustment process is described in the relevant chapter on SA or on SA of HF data.

5.0.1 Overview of Calendar effects in JDemetra

edit : this has a evolved a lot with v3 definition possibilities via GUI and R have to be re described

The following description of the calendar effects in JDemetra+ is strictly based on PALATE, J. (2014).

A natural way for modelling calendar effects consists of distributing the days of each period into different groups. The regression variable corresponding to a type of day (a group) is simply defined by the number of days it contains for each period. Usual classifications are:

- Trading days (7 groups): each day of the week defines a group (Mondays,...,Sundays);
- Working days (2 groups): week days and weekends.

The definition of a group could involve partial days. For instance, we could consider that one half of Saturdays belong to week days and the second half to weekends.

Usually, specific holidays are handled as Sundays and they are included in the group corresponding to "non-working days". This approach assumes that the economic activity on national holidays is the same (or very close to) the level of activity that is typical for Sundays. Alternatively, specific holidays can be considered separately, e.g. by the specification that divided days into three groups:

- Working days (Mondays to Fridays, except for specific holidays),
- Non-working days (Saturdays and Sundays, except for specific holidays),
- Specific holidays.

5.0.2 Summary of the method used in JDemetra+ to compute trading day and working day effects

The computation of trading day and working days effects is performed in four steps:

1. Computation of the number of each weekday performed for all periods.
2. Calculation of the usual contrast variables for trading day and working day.
3. Correction of the contrast variables with specific holidays (for each holiday add +1 to the number of Sundays and subtract 1 from the number of days of the holiday). The correction is not performed if the holiday falls on a Sunday, taking into account the validity period of the holiday.
4. Correction of the constant variables for long term mean effects, > taking into account the validity period of the holiday; see below > for the different cases.

The corrections of the constant variables may receive a weight corresponding to the part of the holiday considered as a Sunday.

An example below illustrates the application of the above algorithm for the hypothetical country in which three holidays are celebrated:

- New Year (a fixed holiday, celebrated on 01 January);
- Shrove Tuesday (a moving holiday, which falls 47 days before Easter Sunday, celebrated until the end of 2012);
- Freedom day (a fixed holiday, celebrated on 25 April).

The consecutive steps in calculation of the calendar for 2012 and 2013 years are explained below.

First, the number of each day of the week in the given month is calculated as it is shown in table below.

Number of each weekday in different months

Month	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Jan-12	5	5	4	4	4	4	5
Feb-12	4	4	5	4	4	4	4
Mar-12	4	4	4	5	5	5	4
Apr-12	5	4	4	4	4	4	5
May-12	4	5	5	5	4	4	4
Jun-12	4	4	4	4	5	5	4
Jul-12	5	5	4	4	4	4	5
Aug-12	4	4	5	5	5	4	4

Month	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Sep-12	4	4	4	4	4	5	5
Oct-12	5	5	5	4	4	4	4
Nov-12	4	4	4	5	5	4	4
Dec-12	5	4	4	4	4	5	5
Jan-13	4	5	5	5	4	4	4
Feb-13	4	4	4	4	4	4	4
Mar-13	4	4	4	4	5	5	5
Apr-13	5	5	4	4	4	4	4
May-13	4	4	5	5	5	4	4
Jun-13	4	4	4	4	4	5	5
Jul-13	5	5	5	4	4	4	4
Aug-13	4	4	4	5	5	5	4
Sep-13	5	4	4	4	4	4	5
Oct-13	4	5	5	5	4	4	4
Nov-13	4	4	4	4	5	5	4
Dec-13	5	5	4	4	4	4	5

Next, the contrast variables are calculated (table below) as a result of the linear transformation applied to the variables presented in table below.

Contrast variables (series corrected for leap year effects)

Month	Mon	Tue	Wed	Thu	Fri	Sat	Length
Jan-12	0	0	-1	-1	-1	-1	0
Feb-12	0	0	1	0	0	0	0.75
Mar-12	0	0	0	1	1	1	0
Apr-12	0	-1	-1	-1	-1	-1	0
May-12	0	1	1	1	0	0	0
Jun-12	0	0	0	0	1	1	0
Jul-12	0	0	-1	-1	-1	-1	0
Aug-12	0	0	1	1	1	0	0
Sep-12	-1	-1	-1	-1	-1	0	0
Oct-12	1	1	1	0	0	0	0
Nov-12	0	0	0	1	1	0	0
Dec-12	0	-1	-1	-1	-1	0	0
Jan-13	0	1	1	1	0	0	0
Feb-13	0	0	0	0	0	0	-0.25
Mar-13	-1	-1	-1	-1	0	0	0
Apr-13	1	1	0	0	0	0	0
May-13	0	0	1	1	1	0	0

Month	Mon	Tue	Wed	Thu	Fri	Sat	Length
Jun-13	-1	-1	-1	-1	-1	0	0
Jul-13	1	1	1	0	0	0	0
Aug-13	0	0	0	1	1	1	0
Sep-13	0	-1	-1	-1	-1	-1	0
Oct-13	0	1	1	1	0	0	0
Nov-13	0	0	0	0	1	1	0
Dec-13	5	5	4	4	4	4	0

In the next step the corrections for holidays is done in the following way:

- New Year: In 2012 it falls on a Sunday. Therefore no correction is applied. In 2013 it falls on a Tuesday. Consequently, the following corrections are applied to the number of each weekday in January: Tuesday -1, Sunday +1, so the following corrections are applied to the contrast variables: -2 for Tuesday and -1 for the other contrast variables.
- Shrove Tuesday: It is a fixed day of the week holiday that always falls on Tuesday. For this reason in 2012 the following corrections are applied to the number of each weekday in February: Tuesday -1, Sunday +1, so the following corrections are applied to the contrast variables: -2 for the contrast variable associated with Tuesday, and -1 for the other contrast variables. The holiday expires at the end of 2012. Therefore no corrections are made for 2013.
- Freedom Day: In 2012 it falls on a Wednesday. Consequently, the following corrections are applied to the number of each weekday in April: Wednesday -1, Sunday +1, so the following corrections are applied to the contrast variables: -2 for Wednesday and -1 for the other contrast variables. In 2013 it falls on Thursday. Therefore, the following corrections are applied to the number of each weekday in April: Thursday -1, Sunday +1, so the following corrections are applied to the contrast variables: -2 for Thursday, and -1 for the other contrast variables.

The result of these corrections is presented in table below.

Contrast variables corrected for holidays

Month	Mon	Tue	Wed	Thu	Fri	Sat	Length
Jan-12	0	0	-1	-1	-1	-1	0
Feb-12	-1	-2	0	-1	-1	-1	0.75
Mar-12	0	0	0	1	1	1	0
Apr-12	-1	-2	-3	-2	-2	-2	0
May-12	0	1	1	1	0	0	0
Jun-12	0	0	0	0	1	1	0
Jul-12	0	0	-1	-1	-1	-1	0

Month	Mon	Tue	Wed	Thu	Fri	Sat	Length
Aug-12	0	0	1	1	1	0	0
Sep-12	-1	-1	-1	-1	-1	0	0
Oct-12	1	1	1	0	0	0	0
Nov-12	0	0	0	1	1	0	0
Dec-12	0	-1	-1	-1	-1	0	0
Jan-13	-1	-1	0	0	-1	-1	0
Feb-13	0	0	0	0	0	0	-0.25
Mar-13	-1	-1	-1	-1	0	0	0
Apr-13	0	0	-1	-2	-1	-1	0
May-13	0	0	1	1	1	0	0
Jun-13	-1	-1	-1	-1	-1	0	0
Jul-13	1	1	1	0	0	0	0
Aug-13	0	0	0	1	1	1	0
Sep-13	0	-1	-1	-1	-1	-1	0
Oct-13	0	1	1	1	0	0	0
Nov-13	0	0	0	0	1	1	0
Dec-13	0	0	-1	-1	-1	-1	0

Finally, the long term corrections are applied on each year of the validity period of the holiday.

- New Year: Correction on the contrasts: +1, to be applied to January of 2012 and 2013.
- Shrove Tuesday: It may fall either in February or in March. It will fall in March if Easter is on or after 17 April. Taking into account the theoretical distribution of Easter, it gives: prob(March) = +0.22147, prob(February) = +0.77853. The correction of the contrasts will be +1.55707 for Tuesday in February 2012 and +0.77853 for the other contrast variables. The correction of the contrasts will be +0.44293 for Tuesday in March 2012, +0.22147 for the other contrast variables.
- Freedom Day: Correction on the contrasts: +1, to be applied to April of 2012 and 2013.

The modifications due to the corrections described above are presented in table below.

Trading day variables corrected for the long term effects

Month	Mon	Tue	Wed	Thu	Fri	Sat	Length
Jan-12	1	1	0	0	0	0	0
Feb-12	-0.22115	-0.44229	0.778853	-0.22115	-0.22115	-0.22115	0.75
Mar-12	0.221147	0.442293	0.221147	1.221147	1.221147	1.221147	0
Apr-12	0	-1	-2	-1	-1	-1	0
May-12	0	1	1	1	0	0	0

Month	Mon	Tue	Wed	Thu	Fri	Sat	Length
Jun-12	0	0	0	0	1	1	0
Jul-12	0	0	-1	-1	-1	-1	0
Aug-12	0	0	1	1	1	0	0
Sep-12	-1	-1	-1	-1	-1	0	0
Oct-12	1	1	1	0	0	0	0
Nov-12	0	0	0	1	1	0	0
Dec-12	0	-1	-1	-1	-1	0	0
Jan-13	0	0	1	1	0	0	0
Feb-13	0	0	0	0	0	0	-0.25
Mar-13	-1	-1	-1	-1	0	0	0
Apr-13	1	1	0	-1	0	0	0
May-13	0	0	1	1	1	0	0
Jun-13	-1	-1	-1	-1	-1	0	0
Jul-13	1	1	1	0	0	0	0
Aug-13	0	0	0	1	1	1	0
Sep-13	0	-1	-1	-1	-1	-1	0
Oct-13	0	1	1	1	0	0	0
Nov-13	0	0	0	0	1	1	0
Dec-13	0	0	-1	-1	-1	-1	0

5.0.3 Mean and seasonal effects of calendar variables

The calendar effects produced by the regression variables that fulfil the definition presented above include a mean effect (i.e. an effect that is independent of the period) and a seasonal effect (i.e. an effect that is dependent of the period and on average it is equal to 0). Such an outcome is inappropriate, as in the usual decomposition of a series the mean effect should be allocated to the trend component and the fixed seasonal effect should be affected to the corresponding component. Therefore, the actual calendar effect should only contain effects that don't belong to the other components.

In the context of JDemetra+ the mean effect and the seasonal effect are long term theoretical effects rather than the effects computed on the time span of the considered series (which should be continuously revised).

The mean effect of a calendar variable is the average number of days in its group. Taking into account that one year has on average 365.25 days, the monthly mean effects for a working days are, as shown in the table below, 21.7411 for week days and 8.696 for weekends.

Monthly mean effects for the Working day variable

Groups of Working day effect	Mean effect
Week days	365.25/12*5/7 = **21.7411**
Weekends	365.25/12*2/7 = **8.696**
Total	365.25/12 = **30.4375**

The number of days by period is highly seasonal, as apart from February, the length of each month is the same every year. For this reason, any set of calendar variables will contain, at least in some variables, a significant seasonal effect, which is defined as the average number of days by period (Januaries..., first quarters...) outside the mean effect. Removing that fixed seasonal effects consists of removing for each period the long term average of days that belong to it. The calculation of a seasonal effect for the working days classification is presented in the table below.

The mean effect and the seasonal effect for the calendar periods

Period	Average number of days	Average number of week days	Mean effect	Seasonal effect
January	31	31*5/7=22.1429	21.7411	0.4018
February	28.25	28.25*5/7=20.1786	21.7411	-1.5625
March	31	31*5/7=22.1429	21.7411	0.4018
April	30	30*5/7=21.4286	21.7411	-0.3125
May	31	31*5/7=22.1429	21.7411	0.4018
June	30	30*5/7=21.4286	21.7411	-0.3125
July	31	31*5/7=22.1429	21.7411	0.4018
August	31	31*5/7=22.1429	21.7411	0.4018
September	30	30*5/7=21.4286	21.7411	-0.3125
October	31	31*5/7=22.1429	21.7411	0.4018
November	30	30*5/7=21.4286	21.7411	-0.3125
December	31	31*5/7=22.1429	21.7411	0.4018
Total	365.25	260.8929	260.8929	0

For a given time span, the actual calendar effect for week days can be easily calculated as the difference between the number of week days in a specific period and the sum of the mean effect and the seasonal effect assigned to this period, as it is shown in the table below for the period 01.2013 – 06.2013.

The calendar effect for the period 01.2013 - 06.2013

Time period (t)	Week days	Mean effect	Seasonal effect	Calendar effect
Jan-2013	23	21.7411	0.4018	0.8571

Time period (t)	Week days	Mean effect	Seasonal effect	Calendar effect
Feb-2013	20	21.7411	-1.5625	-0.1786
Mar-2013	21	21.7411	0.4018	-1.1429
Apr-2013	22	21.7411	-0.3125	0.5714
May-2013	23	21.7411	0.4018	0.8571
Jun-2013	20	21.7411	-0.3125	-1.4286
Jul-2013	23	21.7411	0.4018	0.8571

The distinction between the mean effect and the seasonal effect is usually unnecessary. Those effects can be considered together (simply called mean effects) and be computed by removing from each calendar variable its average number of days by period. These global means effect are considered in the next section.

5.0.4 Impact of the mean effects on the decomposition

When the ARIMA model contains a seasonal difference – something that should always happen with calendar variables – the mean effects contained in the calendar variables are automatically eliminated, so that they don't modify the estimation. The model is indeed estimated on the series/regression variables after differencing. However, they lead to a different linearised series (y_{lin}). The impact of other corrections (mean and/or fixed seasonal) on the decomposition is presented in the next paragraph. Such corrections could be obtained, for instance, by applying other solutions for the long term corrections or by computing them on the time span of the series.

Now the model with ”correct” calendar effects (denoted as C), i.e. effects without mean and fixed seasonal effects, can be considered. To simplify the problem, the model has no other regression effects.

For such a model the following relations hold:

$$y_{\text{lin}} = y - C$$

$$T = F_T\left(y_{\text{lin}}\right)$$

$$S = F_S\left(y_{\text{lin}}\right) + C$$

$$I = F_I\left(y_{\text{lin}}\right)$$

where:

T - the trend;

S - the seasonal component;

I - the irregular component;

F_X - the linear filter for the component X.

Consider next other calendar effects ($\widetilde{C}$) that contain some mean (cm, integrated to the final trend) and fixed seasonal effects (cs, integrated to the final seasonal). The modified equations are now:

$$\widetilde{C} = C + cm + cs$$

$$\tilde{y}_{\text{lin}} = \; y - \widetilde{C} = \; y_{\text{lin}} - cm - cs$$

$$\widetilde{T} = \; F_T\left(\tilde{y}_{\text{lin}}\right) + cm$$

$$\widetilde{S} = \; F_S\left(\tilde{y}_{\text{lin}}\right) + C + cs$$

$$\tilde{I} = \; F_I\left(\tilde{y}_{\text{lin}}\right)$$

Taking into account that F_X is a linear transformation and that[1]

$$F_T\left(\text{cm}\right) = cm$$

$$F_T\left(\text{cs}\right) = 0$$

$$F_S\left(\text{cm}\right) = 0$$

[1] In case of SEATS the properties can be trivially derived from the matrix formulation of signal extraction. They are also valid for X-11 (additive).

$$F_S\left(\mathrm{cs}\right) = cs$$

$$F_I\left(\mathrm{cm}\right) = 0$$

$$F_I\left(\mathrm{cs}\right) = 0$$

The following relationships hold:

$$\widetilde{T} = \ F_T\left(\tilde{y}_{\mathrm{lin}}\right) + cm = F_T\left(y_{\mathrm{lin}}\right) - cm + cm = T$$

$$\widetilde{S} = \ F_S\left(\tilde{y}_{\mathrm{lin}}\right) + C + cs = F_S\left(y_{\mathrm{lin}}\right) - cs + C + cs = S$$

$$\tilde{I} = \ I$$

If we don't take into account the effects and apply the same approach as in the "correct" calendar effects, we will get:

$$\breve{T} = \ F_T\left(\tilde{y}_{\mathrm{lin}}\right) = T - cm$$

$$\breve{S} = \ F_S\left(\tilde{y}_{\mathrm{lin}}\right) + \widetilde{C} = S + cm$$

$$\breve{I} = \ F_I\left(\tilde{y}_{\mathrm{lin}}\right) = I$$

The trend, seasonal and seasonally adjusted series will only differ by a (usually small) constant.

In summary, the decomposition does not depend on the mean and fixed seasonal effects used for the calendar effects, provided that those effects are integrated in the corresponding final components. If these corrections are not taken into account, the main series of the decomposition will only differ by a constant.

5.0.5 Linear transformations of the calendar variables

As far as the RegARIMA and the TRAMO models are considered, any non-degenerated linear transformation of the calendar variables can be used. It will produce the same results (likelihood, residuals, parameters, joint effect of the calendar variables, joint F-test on the coefficients of the calendar variables...). The linearised series that will be further decomposed is invariant to any linear transformation of the calendar variables.

However, it should be mentioned that choices of calendar corrections based on the tests on the individual t statistics are dependent on the transformation, which is rather arbitrary. This is the case in old versions of TRAMO-SEATS. That is why the joint F-test (as in the version of TRAMO-SEATS implemented in TSW+) should be preferred.

An example of a linear transformation is the calculation of the contrast variables. In the case of the usual trading day variables, they are defined by the following transformation: the 6 contrast variables $(\text{No.}\,(\text{Mondays}) - No.\,(\text{Sundays})\,, \ldots No.\,(\text{Saturdays}) - No.(Sundays))$ used with the length of period.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \text{Mon} \\ \text{Tue} \\ \text{Wed} \\ \text{Thu} \\ \text{Fri} \\ \text{Sat} \\ \text{Sun} \end{bmatrix} = \begin{bmatrix} Mon - Sun \\ Tue - Sun \\ Wed - Sun \\ Thu - Sun \\ Fri - Sun \\ Sat - Sun \\ \text{Length of period} \end{bmatrix}$$

For the usual working day variables, two variables are used: one contrast variable and the length of period

$$\begin{bmatrix} 1 & -\frac{5}{2} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \text{Week} \\ \text{Weekend} \end{bmatrix} = \begin{bmatrix} \text{Contrast week} \\ \text{Length of period} \end{bmatrix}$$

The Length of period variable is defined as a deviation from the length of the month (in days) and the average month length, which is equal to 30.4375. Instead, the leap-year variable can be used here (see Regression sections in RegARIMA or Tramo)[2].

Such transformations have several advantages. They suppress from the contrast variables the mean and the seasonal effects, which are concentrated in the last variable. So, they lead to fewer correlated variables, which are more appropriate to be included in the regression model. The sum of the effects of each day of the week estimated with the trading (working) day contrast variables cancel out.

[2] GÓMEZ, V., and MARAVALL, A (2001b).

5.0.6 Handling of specific holidays

check vs GUI (v3) and rjd3 modelling

Three types of holidays are implemented in JDemetra+:

- Fixed days, corresponding to the fixed dates in the year (e.g. New Year, Christmas).
- Easter related days, corresponding to the days that are defined in relation to Easter (e.g. Easter +/- n days; example: Ascension, Pentecost).
- Fixed week days, corresponding to the fixed days in a given week of a given month (e.g. Labor Day celebrated in the USA on the first Monday of September).

From a conceptual point of view, specific holidays are handled in exactly the same way as the other days. It should be decided, however, to which group of days they belong. Usually they are handled as Sundays. This convention is also used in JDemetra+. Therefore, except if the holiday falls on a Sunday, the appearance of a holiday leads to correction in two groups, i.e. in the group that contains the weekday, in which holiday falls, and the group that contains the Sundays.

Country specific holidays have an impact on the mean and the seasonal effects of calendar effects. Therefore, the appropriate corrections to the number of particular days (which are usually the basis for the definition of other calendar variables) should be applied, following the kind of holidays. These corrections are applied to the period(s) that may contain the holiday. The long term corrections in JDemetra+ don't take into account the fact that some moving holidays could fall on the same day (for instance the May Day and the Ascension). However, those events are exceptional, and their impact on the final result is usually not significant.

5.0.6.1 Fixed day

The probability that the holiday falls on a given day of the week is 1/7. Therefore, the probability to have 1 day more that is treated like Sunday is 6/7. The effect on the means for the period that contains the fixed day is presented in the table below (the correction on the calendar effect has the opposite sign).

The effect of the fixed holiday on the period, in which it occurred

Sundays	Others days	Contrast variables
+ 6/7	- 1/7	1/7 - (+ 6/7)= -1

5.0.6.2 Easter related days

Easter related days always fall the same week day (denoted as Y in the table below: The effects of the Easter Sunday on the seasonal means). However, they can fall during different periods (months or quarters). Suppose that, taking into account the distribution of the dates for Easter and the fact that this holiday falls in one of two periods, the probability that Easter falls during the period m is p, which implies that the probability that it falls in the period $m+1$ is $1-p$. The effects of Easter on the seasonal means are presented in the table below.

The effects of the Easter Sunday on the seasonal means

|**Period | Sundays Days X Others days Contrast Y Other contrasts** |———| —— ——- ——— ———— ————- ————— |m |+ p - p 0 - 2p - p |m+1 | + (1-p) - (1-p) 0 - 2×(1-p) - (1-p)

The distribution of the dates for Easter may be approximated in different ways. One of the solutions consists of using some well-known algorithms for computing Easter on a very long period. JDemetra+ provides the Meeus/Jones/Butcher's and the Ron Mallen's algorithms (they are identical till year 4100, but they slightly differ after that date). Another approach consists in deriving a raw theoretical distribution based on the definition of Easter. It is the solution used for Easter related days. It is shortly explained below.

The date of Easter in the given year is the first Sunday after the full moon (the Paschal Full Moon) following the northern hemisphere's vernal equinox. The definition is influenced by the Christian tradition, according to which the equinox is reckoned to be on 21 March[3] and the full moon is not necessarily the astronomically correct date. However, when the full moon falls on Sunday, then Easter is delayed by one week. With this definition, the date of Easter Sunday varies between 22 March and 25 April. Taking into account that an average lunar month is 29.530595 days the approximated distribution of Easter can be derived. These calculations do not take into account the actual ecclesiastical moon calendar.

For example, the probability that Easter Sunday falls on 25 March is 0.004838 and results from the facts that the probability that 25 March falls on a Sunday is 1/7 and the probability that the full moon is on 21 March, 22 March, 23 March or 24 March is 5/29.53059. The probability that Easter falls on 24 April is 0.01708 and results from the fact that the probability that 24 April is Sunday is 1/7 and takes into account that 18 April is the last acceptable date for the full moon. Therefore the probability that the full moon is on 16 April or 17 April is 1/29.53059 and the probability that the full moon is on 18 April is 1.53059/29.53059.

The approximated distribution of Easter dates

Day	Probability
22 March	1/7 * 1/29.53059
23 March	1/7 * 2/29.53059

[3] In fact, astronomical observations show that the equinox occurs on 20 March in most years.

Day	Probability
24 March	1/7 * 3/29.53059
25 March	1/7 * 4/29.53059
26 March	1/7 * 5/29.53059
27 March	1/7 * 6/29.53059
28 March	1/29.53059
29 March	1/29.53059
...	...
18 April	1/29.53059
19 April	1/7 * (6 + 1.53059)/29.53059
20 April	1/7 * (5 + 1.53059)/29.53059
21 April	1/7 * (4 + 1.53059)/29.53059
22 April	1/7 * (3 + 1.53059)/29.53059
23 April	1/7 * (2 + 1.53059)/29.53059
24 April	1/7 * (1 + 1.53059)/29.53059
25 April	1/7 * 1.53059/29.53059

5.0.6.3 Fixed week days

Fixed week days always fall on the same week day (denoted as Y in the table below) and in the same period. Their effect on the seasonal means is presented in the table below.

The effect of the fixed week holiday on the period, in which it occurred

Sundays	Day Y	Others days
+ 1	- 1	0

The impact of fixed week days on the regression variables is zero because the effect itself is compensated by the correction for the mean effect.

5.0.7 Holidays with a validity period

When a holiday is valid only for a given time span, JDemetra+ applies the long term mean corrections only on the corresponding period. However, those corrections are computed in the same way as in the general case.

It is important to note that using or not using mean corrections will impact in the estimation of the RegARIMA and TRAMO models. Indeed, the mean corrections do not disappear after differencing. The differences between the SA series computed with or without mean corrections will no longer be constant.

5.0.8 Different Kinds of calendars

see link with GUI

This scenario presents how to define different kinds of calendars. These calendars can be applied to the specifications that take into account country-specific holidays and can be used for detecting and estimating the calendar effects.

The calendar effects are those parts of the movements in the time series that are caused by different number of weekdays in calendar months (or quarters). They arise as the number of occurrences of each day of the week in a month (or a quarter) differs from year to year. These differences cause regular effects in some series. In particular, such variation is caused by a leap year effect because of an extra day inserted into February every four years. As with seasonal effects, it is desirable to estimate and remove calendar effects from the time series.

The calendar effects can be divided into a mean effect, a seasonal part and a structural part. The mean effect is independent from the period and therefore should be allocated to the trend-cycle. The seasonal part arises from the properties of the calendar that recur each year. For one thing, the number of working days of months with 31 calendar days is on average larger than that of months with 30 calendar days. This effect is part of the seasonal pattern captured by the seasonal component (with the exception of leap year effects). The structural part of the calendar effect remains to be determined by the calendar adjustment. For example, the number of working days of the same month in different years varies from year to year.

Both X-12-ARIMA/X-13ARIMA-SEATS and TRAMO/SEATS estimate calendar effects by adding some regressors to the equation estimated in the pre-processing part (RegARIMA or TRAMO, respectively). Regressors mentioned above are generated from the default calendar or the user defined calendar.

The calendars of JDemetra+ simply correspond to the usual trading days contrast variables based on the Gregorian calendar, modified to take into account some specific holidays. Those holidays are handled as "Sundays" and the variables are properly adjusted to take into account the long term mean effects.

5.0.9 Tests for residual trading days

We consider below tests on the seasonally adjusted series (sa_t) or on the irregular component (irr_t). When the reasoning applies on both components, we will use y_t. The functions *stdev* stands for "standard deviation" and *rms* for "root mean squares"

The tests are computed on the log-transformed components in the case of multiplicative decomposition.

TD are the usual contrasts of trading days, 6 variables (no specific calendar).

5.0.9.1 Non significant irregular

When irr_t is not significant, we don't compute the test on it, to avoid irrelevant results. We consider that irr_t is significant if $stdev(irr_t) > 0.01$ (multiplicative case) or if $stdev(irr_t)/rms(sa_t) > 0.01$ (additive case).

5.0.9.2 F test

The test is the usual joint F-test on the TD coefficients, computed on the following models:

5.0.9.2.1 Autoregressive model (AR modelling option)

We compute by OLS:

$$y_t = \mu + \alpha y_{t-1} + \beta TD_t + \epsilon_t$$

5.0.9.2.2 Difference model

We compute by OLS:

$$\Delta y_t - \overline{\Delta y_t} = \beta TD_t + \epsilon_t$$

So, the latter model is a restriction of the first one ($\alpha = 1, \mu = \mu = \overline{\Delta y_t}$)

The tests are the usual joint F-tests on $\beta \quad (H_0 : \beta = 0)$.

By default, we compute the tests on the 8 last years of the components, so that they might highlight moving calendar effects.

Remark:

In Tramo, a similar test is computed on the residuals of the Arima model. More exactly, the F-test is computed on $e_t = \beta TD_t + \epsilon_t$, where e_t are the one-step-ahead forecast errors.

6 Algorithms for benchmarking and temporal disagreggation

In this chapter we describe the practical implementation, the underlying theory in a dedicated chapter.(link)

6.1 Benchmarking overview

Often one has two (or multiple) datasets of different frequency for the same target variable. Sometimes, however, these data sets are not coherent in the sense that they don't match up. Benchmarking[^1] is a method to deal with this situation. An aggregate of a higher-frequency measurement variables is not necessarily equal to the corresponding lower-frequency less-aggregated measurement. Moreover, the sources of data may have different reliability levels. Usually, less frequent data are considered more trustworthy as they are based on larger samples and compiled more precisely. The more reliable measurements, hence often the less frequent, will serve as benchmark.

In seasonal adjustment methods benchmarking is the procedure that ensures the consistency over the year between adjusted and non-seasonally adjusted data. It should be noted that the [ESS Guidelines on Seasonal Adjustment (2015)] (https://ec.europa.eu/eurostat/documents/3859598/6830795/K GQ-15-001-EN-N.pdf/d8f1e5f5-251b-4a69-93e3-079031b74bd3), do not recommend benchmarking as it introduces a bias in the seasonally adjusted data. The U.S. Census Bureau also points out that "*forcing the seasonal adjustment totals to be the same as the original series annual totals can degrade the quality of the seasonal adjustment, especially when the seasonal pattern is undergoing change. It is not natural if trading day adjustment is performed because the aggregate trading day effect over a year is variable and moderately different from zero*"[^2]. Nevertheless, some users may need that the annual totals of the seasonally adjusted series match the annual totals of the original, non-seasonally adjusted series[^3].

According to the [ESS Guidelines on Seasonal Adjustment (2015)] (https://ec.europa.eu/eurostat/documents/385 GQ-15-001-EN-N.pdf/d8f1e5f5-251b-4a69-93e3-079031b74bd3), the only benefit of this approach is that there is consistency over the year between adjusted and the non-seasonally adjusted data; this can be of particular interest when low-frequency (e.g. annual) benchmarking figures officially exist (e.g. National Accounts, Balance of Payments, External Trade, etc.) and where users' needs for time consistency are stronger.

6.2 Tools

6.2.1 Benchmarking with GUI

1. With the pre-defined specifcations the benchmarking functionality is not applied by default following the *ESS Guidelines on Seasonal Adjustment* (2015) recommendations. It means that once the user has seasonally adjustd the series with a pre-defined specifcation the *Benchmarking* node is empty. To execute benchmarking click on the *Specifications* button and activate the checkbox in the *Benchmarking* section.

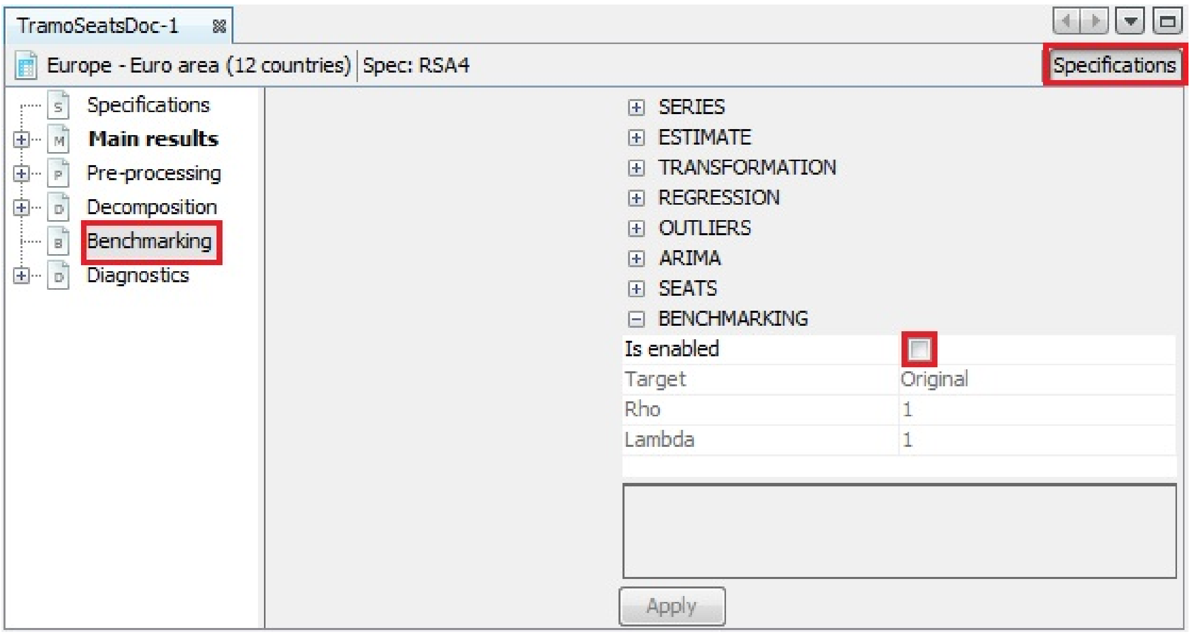

Figure 6.1: Text

Benchmarking option – a default view

2. Three parameters can be set here. *Target* specifies the target variable for the benchmarking procedure. It can be either the *Original* (the raw time series) or the *Calendar Adjusted* (the time series adjusted for calendar effects). *Rho* is a value of the AR(1) parameter (set between 0 and 1). By default it is set to 1. Finally, *Lambda* is a parameter that relates to the weights in the regression equation. It is typically equal to 0 (for an additive decomposition), 0.5 (for a proportional decomposition) or 1 (for a multiplicative decomposition). The default value is 1.

3. To launch the benchmarking procedure click on the **Apply** button. The results are displayed in four panels. The top-left one compares the original output from the seasonal adjustment procedure with the result from applying a benchmarking to the seasonal

adjustment. The bottom-left panel highlights the differences between these two results. The outcomes are also presented in a table in the top-right panel. The relevant statistics concerning relative differences are presented in the bottom-right panel.

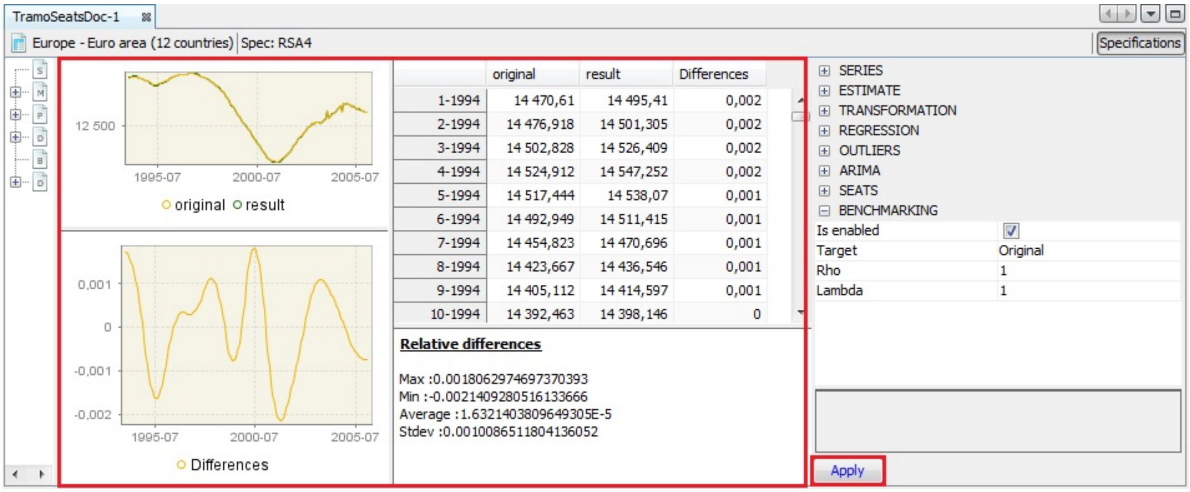

Figure 6.2: Text

The results of the benchmarking procedure

4. Both pictures and the table can be copied the usual way (see the *Simple seasonal adjustment of a single time series* scenario).

Options for benchmarking results

5. To export the result of the benchmarking procedure (*benchmarking.result*) and the target data (*benchmarking.target*) one needs to once execute the seasonal adjustment with benchmarking using the muli-processing option (see the *Simple seasonal adjustment of multiple time series* scenario. Once the muli-processing is executed, select the *Output* item from the *SAProcessing* menu.

The *SAProcessing* menu

6. Expand the "+" menu and choose an appropriate data format (here Excel has been chosen). It is possible to save the results in TXT, XLS, CSV, and CSV matrix formats. Note that the available content of the output depends on the output type.

Exporting data to an Excel file

7. Chose the output items that refer to the results from the benchmarking procedure, move them to the window on the right and click **OK**.

Exporting the results of the benchmarking procedure

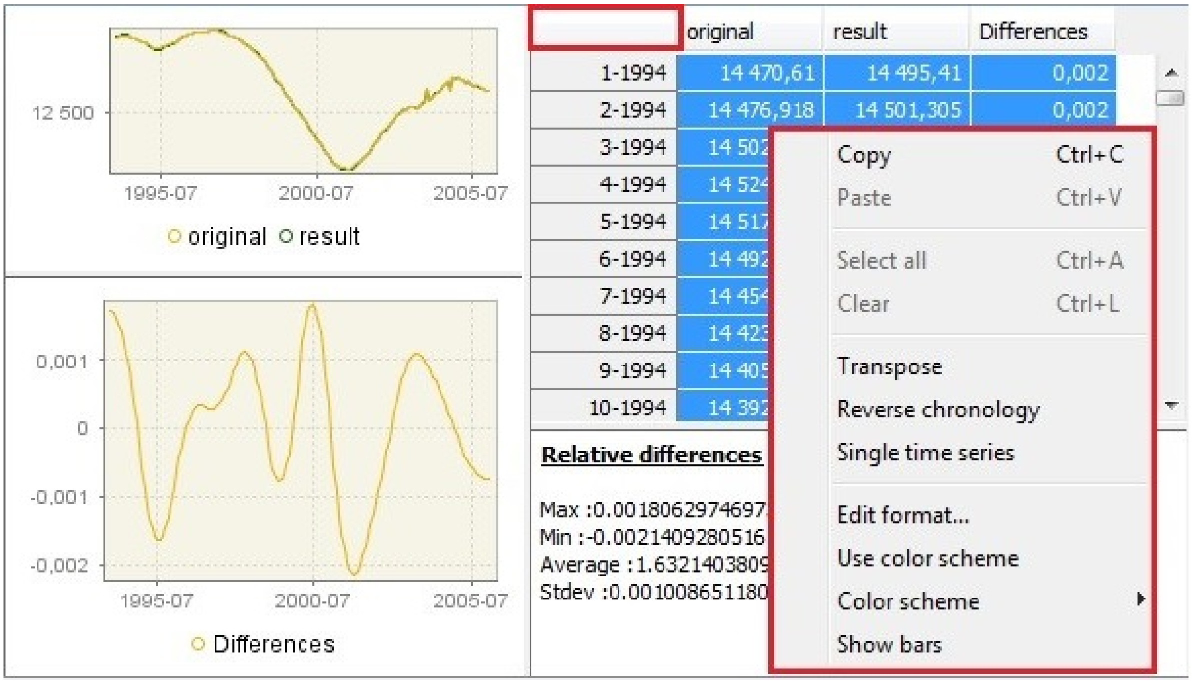

Figure 6.3: Text

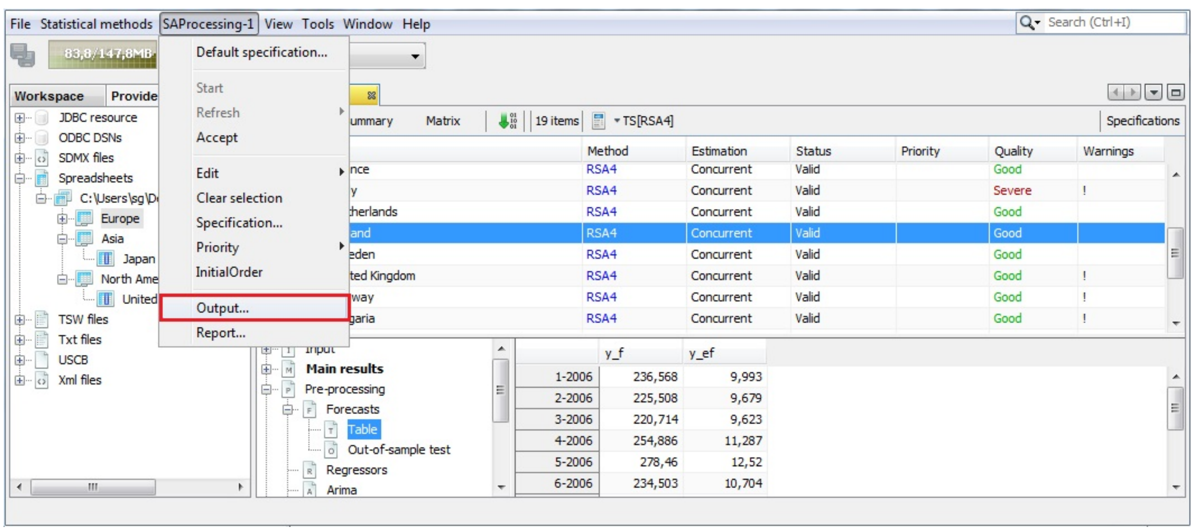

Figure 6.4: Text

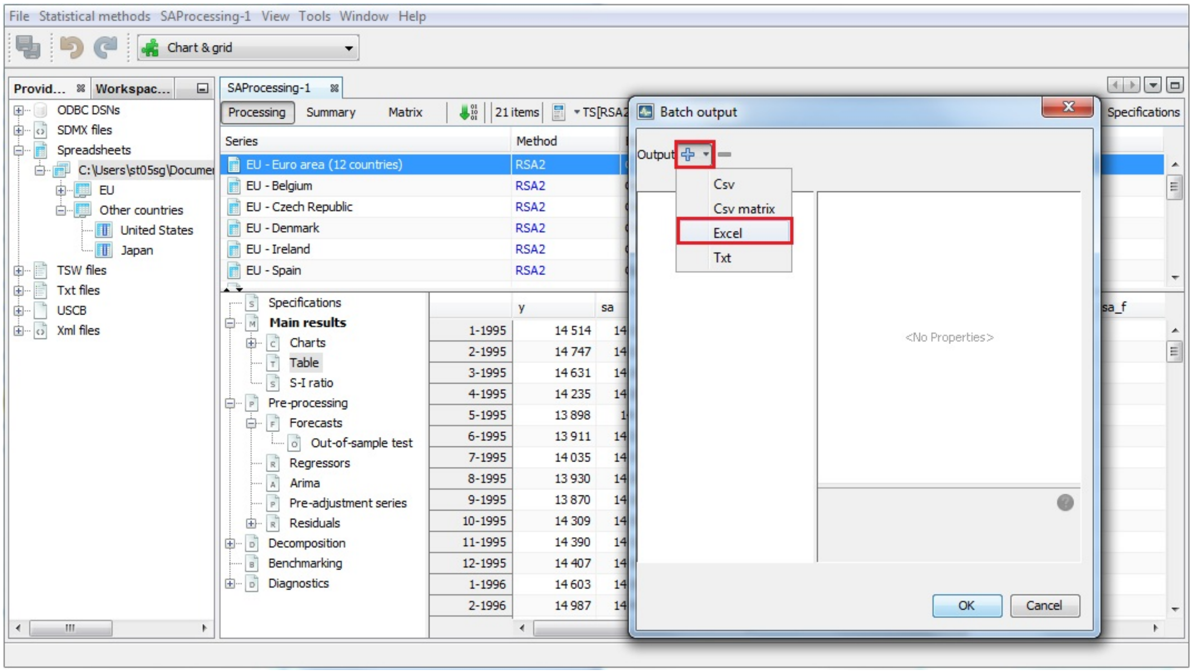

Figure 6.5: Text

6.2.2 Benchmarking in R

package rjd3bench orga doc - here - in package - example

6.3 References

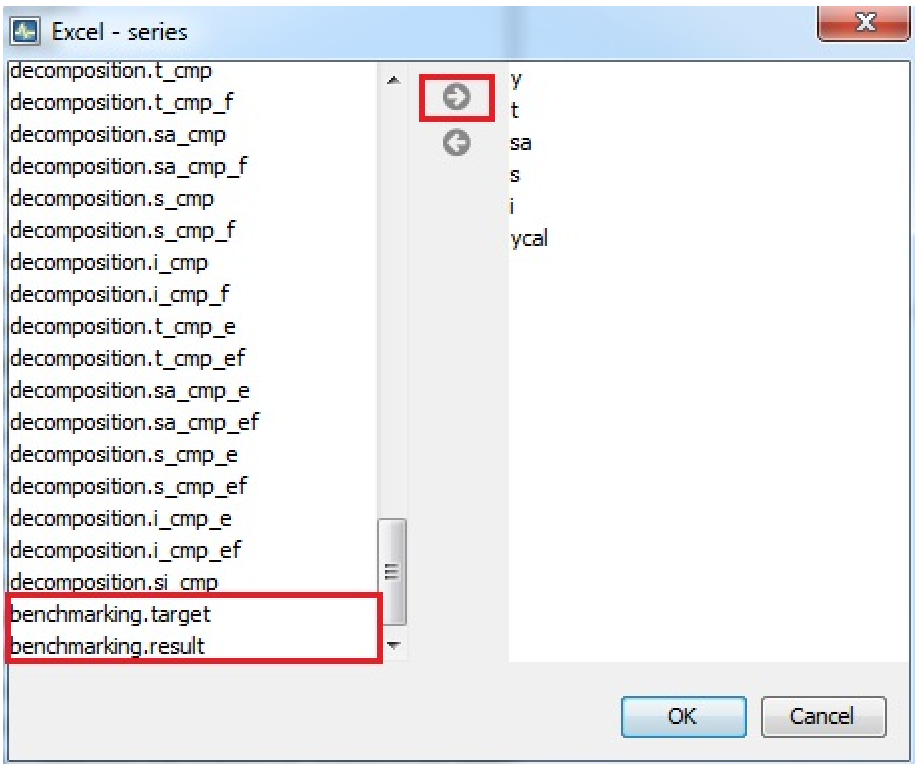

Figure 6.6: Text

7 Trend-cycle estimation

7.1 Motivation

7.2 Estimation Methods

7.3 Tools

7.3.1 rjd3 highfeq package

7.3.2 rjdfilters package

8 Nowcasting

8.1 Motivation

Underlying Theory: references ?

8.2 Tools

- plug in ?
- R package ?

9 Graphical User Interface

9.1 Overview

10 R packages

10.1 Available algorithms

table

10.2 Organisation overview

a suite (order)

general output organisation

10.3 Installation procedure

10.4 Interaction with GUI

10.5 Full list

10.5.1 rjd3modelling

main functions: table

11 Plug-ins for JDemetra+

11.1 Main functions

table

12 Production issues and cruncher use

12.0.1 Revision Policies

13 Tool selection issues and heuristics

14 Spectral Analysis Principles and Tools

14.1 Spectral analysis concepts

A time series x_t with stationary covariance, mean μ and k^{th} autocovariance $E(x_t - \mu)(x_{t-k}\mu)) = \gamma(k)$ can be described as a weighted sum of periodic trigonometric functions: $sin(\omega t)$ and $cos(\omega t)$, where $\omega = \frac{2*pi}{T}$ denotes frequency. Spectral analysis investigates this frequency domain representation of x_t to determine how important cycles of different frequencies are in accounting for the behavior of x_t.

Assuming that the autocovariances $\gamma(k)$ are absolutely summable($\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$), the autocovariance generating function, which summarizes these autocovariances through a scalar valued function, is given by equation [1][1].

$acgf(z) = \sum_{k=-\infty}^{\infty} z^k \gamma(k)$,

where z denotes complex scalar.

Once the equation [1]is divided by π and evaluated at some $z = e^{-i\omega} = cos\omega - isin\omega$, where $i = \sqrt{-1}$ and ω is a real scalar, $-\infty < \omega < \infty$, the result of this transformation is called a population spectrum $f(\omega)$for x_t, given in equation [2][2].

$$f(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \gamma(k)$$

Therefore, the analysis of the population spectrum in the frequency domain is equivalent to the examination of the autocovariance function in the time domain analysis; however it provides an alternative way of inspecting the process. Because $f(\omega)$d is interpreted as a contribution to the variance of components with frequencies in the range $(\omega, \ \omega + d\omega)$, a peak in the spectrum indicates an important contribution to the variance at frequencies near the value that corresponds to this peak.

As $e - i\omega = cos\omega - isin\omega$, the spectrum can be also expressed as in equation [3].

$$f(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} (cos\omega k - isin\omega k)\gamma(k)$$

[1] HAMILTON, J.D. (1994).
[2] HAMILTON, J.D. (1994).

Since $\gamma(k) = \gamma(-k)$ (i.e. $\gamma(k)$ is an even function of k) and $\sin(-x) =$ -$\sin x$, [3] can be presented as equation

$$f(\omega) = \frac{1}{\pi}[\ \gamma(0) + 2\sum_{k=1}^{\infty}\ \gamma(k)cos\ \text{k}]$$

,

This implies that if autocovariances are absolutely summable the population spectrum exists and is a continuous, real-valued function of ω. Due to the properties of trigonometric functions ($\cos(-\omega k) = \cos($ k) and $\cos(\omega + 2\pi j)k = cos(\omega k)$) the spectrum is a periodic, even function of ω, symmetric around $\omega = 0$. Therefore, the analysis of the spectrum can be reduced to the interval $(-\pi, \pi)$. The spectrum is non-negative for all $\omega \in (-\pi, \pi)$.

The shortest cycle that can be distinguished in a time series lasts two periods. The frequency which corresponds to this cycle is $\omega = \pi$ and is called the Nyquist frequency. The frequency of the longest cycles that can be observed in the time series with n observations is $\omega = \frac{2\pi}{n}$ and is called the fundamental (Fourier) frequency.

Note that if x_t is a white noise process with zero mean and variance σ^2, then for all $|k| > 0$ $\gamma(k) = 0$ and the spectrum of x_t is constant ($f(\omega) = \frac{\sigma^2}{\pi}$) since each frequency in the spectrum contributes equally to the variance of the process[3].

The aim of spectral analysis is to determine how important cycles of different frequencies are in accounting for the behaviour of a time series[4]. Since spectral analysis can be used to detect the presence of periodic components, it is a natural diagnostic tool for detecting trading day effects as well as seasonal effects[5]. Among the tools used for spectral analysis are the autoregressive spectrum and the periodogram.

The explanations given in the subsections of this node derive mainly from DE ANTONIO, D., and PALATE, J. (2015) and BROCKWELL, P.J., and DAVIS, R.A. (2006).

comment1: end old intro: ok

14.1.1 Theoretical spectral density of an ARIMA model

14.2 Spectral density estimation

14.2.1 Method 1: The periodogram

For any given frequency ω the sample periodogram is the sample analog of the sample spectrum. In general, the periodogram is used to identify the periodic components of unknown

[3]BROCKWELL, P.J., and DAVIS, R.A. (2002).
[4]HAMILTON, J.D. (1994).
[5]SOKUP, R.J., and FINDLEY, D. F. (1999).

frequency in the time series. X-13ARIMA-SEATS and TRAMO-SEATS use this tool for detecting seasonality in raw time series and seasonally adjusted series. Apart from this it is applied for checking randomness of the residuals from the ARIMA model.

To define the periodogram, first consider the vector of complex numbers[6]:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix} \in \mathbb{C}^n$$

where $\mathbb{C}^n$ is the set of all column vectors with complex-valued components.

The Fourier frequencies associated with the sample size n are defined as a set of values $\omega_j = \frac{2\pi j}{n}$, $j = -[\frac{n-1}{2}], \dots, [\frac{n}{2}]$, $-\pi < \omega_j \leq \pi$, $j \in F_n$, where $[n]$ denotes the largest integer less than or equal to n. The Fourier frequencies, which are called harmonics, are given by integer multiples of the fundamental frequency $\frac{2\pi}{n}$.

Now the n vectors $e_j = n^{-\frac{1}{2}}(e^{-i\omega_j}, e^{-i2\omega_j}, \dots, e^{-in\omega_j})'$ can be defined. Vectors $e_1, \dots, e_n$ are orthonormal in the sense that:

$$\mathbf{e}_j^* \mathbf{e}_k = n^{-1} \sum_{r=1}^{n} e^{ir(\omega_j - \omega_k)} = \begin{matrix} 1, \; if \; j = k \\ 0, \; if \; j \neq k \end{matrix}$$

where $\mathbf{e}_j^*$ denotes the row vector, which k^{th} component is the complex conjugate of the k^{th} component of $\mathbf{e}_j$.[7] These vectors are a basis of F_n, so that any $\mathbf{x} \in \mathbb{C}^n$ can be expressed as a sum of n components:

$$\mathbf{x} = \sum_{j=-[\frac{n-1}{2}]}^{[\frac{n}{2}]} a_j \mathbf{e}_j$$

where the coefficients $a_j = \mathbf{e}_j^* \mathbf{x} = n^{-\frac{1}{2}} \sum_{t=1}^{n} x_t e^{-it\omega_j}$ are derived from [3] by multiplying the equation on the left by $\mathbf{e}_j^*$ and using [1].

The sequence of $\{a_j, j \in F_n\}$ is referred as a discrete Fourier transform of $\mathbf{x} \in \mathbb{C}^n$ and the periodogram $I(\omega_j)$ of $\mathbf{x}$ at Fourier frequency $\omega_j = \frac{2\pi j}{n}$ is defined as the square of the Fourier transform $\{a_j\}$ of $\mathbf{x}$:

[6] BROCKWELL, P.J., and DAVIS, R.A. (2002).

[7] For details see BROCKWELL, P.J., and DAVIS, R.A. (2006).

$$I(\omega_j)=|a_j|^2 \ = n^{-\ 1}|\sum_{t=1}^{n} x_t e^{-it\omega_j}|^2$$

From [2] and [3] it can be shown that in fact the periodogram decomposes the total sum of squares $\sum_{t=1}^{n} |x_t|^2$ into a sums of components associated with the Fourier frequencies

$$\omega_j$$

:

$$\sum_{t=1}^{n} |x_t|^2 = \sum_{j=-[\frac{n-1}{2}]}^{[\frac{n}{2}]} |a_j|^2 = \sum_{j=-[\frac{n-1}{2}]}^{[\frac{n}{2}]} I(\omega_j)$$

If $\mathbf{x} \in R^n$, ω_j and $-\omega_j$ are both in $[-\pi, -\pi]$ and a_j is presented in its polar form (i.e. $a_j = r_j \exp(i\theta_j)$), where r_j is the modulus of a_j, then [3] can be rewritten in the form:

$$\mathbf{x} = a_0\mathbf{e}_0 + \sum_{j=1}^{[\frac{n-1}{2}]} 2^{1/2} r_j(\mathbf{c}_j \cos\theta_j - \mathbf{s}_j \sin\theta_j) + a_{n/2}\mathbf{e}_{n/2}$$

The orthonormal basis for R^n is $\{\mathbf{e}_0, \mathbf{c}_1, \mathbf{s}_1, \dots, \mathbf{c}_{[\frac{n-1}{2}]}, \mathbf{s}_{[\frac{n-1}{2}]}, \mathbf{e}_{\frac{n}{2}(excluded\ if\ n\ is\ odd)}\}$, where:

$\mathbf{e}_0$ is a vector composed of n elements equal to $n^{-1/2}$, which implies that $\mathbf{a}_0\mathbf{e}_0 = (n^{-1}\sum_{t=1}^{n} x_t, \dots, n^{-1}\sum_{t=1}^{n} x_t)^{'}$;

$$\mathbf{c}_j = (\frac{n}{2})^{-1/2}(\cos\omega_j, \cos 2\omega_j, \dots, \cos n\omega_j)^{'}, for 1 \leq j \leq [\frac{(n-1)}{2}]$$

;

$$\mathbf{s}_j = (\frac{n}{2})^{-1/2}(\sin\omega_j, \sin 2\omega_j, \dots, \sin n\omega_j)^{'},\ for\ 1 \leq j \leq [\frac{(n-1)}{2}]$$

;

$$\mathbf{e}_{n/2} = (-(n^{-\frac{1}{2}}), n^{-\frac{1}{2}}, \dots, -(n)^{-\frac{1}{2}}), n^{-\frac{1}{2}})^{'}$$

.

Equation [5] can be seen as an OLS regression of x_t on a constant and the trigonometric terms. As the vector of explanatory variables includes n elements, the number of explanatory variables in [5] is equal to the number of observations. HAMILTON, J.D. (1994) shows that the explanatory variables are linearly independent, which implies that an OLS regression yields a perfect fit (i.e. without an error term). The coefficients have the form of a simple OLS projection of the data on the orthonormal basis:

$$\hat{a}_0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t$$

[7]

$$\hat{a}_{n/2} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (-1)^t x_t (\text{only when n is even})$$

[8]

$$\hat{a}_0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t$$

[9]

$$\hat{\alpha}_j = 2^{1/2} r_j \cos\theta_j = (\frac{n}{2})^{-1/2} \sum_{t=1}^{n} x_t \cos(t\frac{2\pi j}{n}), j = 1, \dots, [\frac{n-1}{2}]$$

[10]

$$\hat{\beta}_j = 2^{1/2} r_j \sin\theta_j = (\frac{n}{2})^{-1/2} \sum_{t=1}^{n} x_t \sin(t\frac{2\pi j}{n}), j = 1, \dots, [\frac{n-1}{2}]$$

[11]

With [5] the total sum of squares $\sum_{t=1}^{n} |x_t|^2$ can be decomposed into $2 \times [\frac{n-1}{2}]$ components corresponding to $\mathbf{c}_j$ and $\mathbf{s}_j$, which are grouped to produce the "frequency ω_j" component for $1 \geq j \geq [\frac{n-1}{2}]$. As it is shown in the table below, the value of the periodogram at the frequency ω_j is the contribution of the jth harmonic to the total sum of squares $\sum_{t=1}^{n} |x_t|^2$.

Decomposition of sum of squares into components corresponding to the harmonics

{: .table .table-style}

Frequency	**Degrees of freedom**	**Sum of squares decomposition**
ω_0(mean)	1	$a_0^2 = n^{-1}(\sum_{t=1}^{n} x_t)^2 = I(0)$
ω_1	2	$2r_1^2 = 2\|a_1\|^2 = 2I(\omega_1)$
$\vdots$	$\vdots$	$\vdots$
ω_k	2	$2r_k^2 = 2\|a_k\|^2 = 2I(\omega_k)$
$\vdots$	$\vdots$	$\vdots$
$\omega_{n/2} = \pi$ (excluded if n is odd)	1	$a_{n/2}^2 = I(\pi)$
Total	**n**	$\mathbf{\sum_{t=1}^{n} x_t^2}$

Source: DE ANTONIO, D., and PALATE, J. (2015).

Obviously, if series were random then each component $I(\omega\_j)$ would have the same expectation. On the contrary, when the series contains a systematic sine component having a frequency j and amplitude A then the sum of squares $I(\omega_j)$ increases with A. In practice, it is unlikely that the frequency j of an unknown systematic sine component would exacly match any of the frequencies, for which peridogram have been calcuated. Therefore, the periodogram would show an increase in intensities in the immediate vicinity of j.[8]

Note that in JDemetra+ the periodogram object corresponds exactly to the contribution to the sum of squares of the standardised data, since the series are divided by their standard deviation for computational reasons.

Using the decomposition presented in table above the periodogram can be expressed as:

$$I(\omega_j){=}r_j^2 = \frac{1}{2}(\alpha_j^2 + \beta_j^2) = \ \frac{1}{n}(\sum_{t=1}^{n} x_t \cos\left(t\frac{2\pi j}{n}\ \right)\ )^2 + \frac{1}{n}(\sum_{t=1}^{n} x_t \sin(t\frac{2\pi j}{n})\ )^2$$

[12]

where $j = 0, \ldots, lbrack\frac{n}{2}rbrack.$

Since $\mathbf{x} - \overline{\mathbf{x}}$ are generated by an orthonormal basis, and $\overline{\mathbf{x}}{=}a_0\mathbf{e}_0$ [5] can be rearranged to show that the sum of squares is equal to the sum of the squared coefficients:

$$\mathbf{x} - a_0\mathbf{e}_0 = \sum_{j=1}^{[(n-1)/2]}(\alpha_j\mathbf{c}_j + \beta_j\mathbf{s}_j) + a_{n/2}\mathbf{e}_{n/2}$$

. [13]

Thus the sample variance of

$$x_t$$

can be expressed as:

$$n^{-1}\sum_{t=1}^{n}\left(x_t - \overline{x}\right)^2 = n^{-1}(\sum_{k=1}^{[(n-1)/2]} 2{r_j}^2 + {a_{n/2}}^2)$$

, [14]

where $a_{n/2}^2$ is excluded if n is odd.

The term $2{r_j}^2$ in [14] is then the contribution of the j^{th} harmonic to the variance and [14] shows then how the total variance is partitioned.

The periodogram ordinate $I(\omega_j)$ and the autocovariance coefficient $\gamma(k)$ are both quadratic forms of x_t. It can be shown that the periodogram and autocovarinace function are related

[8]BOX, G.E.P., JENKINS, G.M., and REINSEL, G.C. (2007).

and the periodogram can be written in terms of the sample autocovariance function for any non-zero Fourier frequency ω_j :[9]

$$I(\omega_j) = \sum_{|k|<n} \hat{\gamma}(k)\ e^{-ik\omega_j} = \hat{\gamma}(0)\ + 2\sum_{k=1}^{n-1} \hat{\gamma}(k) \cos{(k\omega_j)}$$

and for the zero frequency $I(0) = n|\overline{x}|^2$.

Once comparing [15] with an expression for the spectral density of a stationary process:

$$f(\omega\ ) = \frac{1}{2\pi} \sum_{k<-\infty}^{\infty} \gamma(k)\ e^{-ik\omega}\ \ = \frac{1}{2\pi} lbrack \gamma(0)\ + 2(\sum_{k=1}^{\infty} \gamma(k) \cos{(k\omega\ )}) rbrack$$

It can be noticed that the periodogram is a sample analog of the population spectrum. In fact, it can be shown that the periodogram is asymptotically unbiased but inconsistent estimator of the population spectrum $f(\omega)$.[^75] Therefore, the periodogram is a wildly fluctuating, with high variance, estimate of the spectrum. However, the consistent estimator can be achieved by applying the different linear smoothing filters to the periodogram, called lag-window estimators. The lag-window estimators implemented in JDemetra+ includes square, Welch, Tukey, Barlett, Hanning and Parzen. They are described in DE ANTONIO, D., and PALATE, J. (2015). Alternatively, the model-based consistent estimation procedure, resulting in autoregressive spectrum estimator, can be applied.

comment2: end part theory>spectral analysis>periodogram

14.2.2 Method 2: Autoregressive spectrum estimation

BROCKWELL, P.J., and DAVIS, R.A. (2006) point out that for any real-valued stationary process (x_t) with continuous spectral density $f(\omega)$ it is possible to find both $AR(p)$ and $MA(q)$ processes which spectral densities are arbitrarily close to $f(\omega)$. For this reason, in some sense, (x_t) can be approximated by either $AR(p)$ or $MA(q)$ process. This fact is a basis of one of the methods of achieving a consistent estimator of the spectrum, which is called an autoregressive spectrum estimation. It is based on the approximation of the stochastic process (x_t) by an autoregressive process of sufficiently high order p:

$$x_t = \mu + (\phi_1 B + ... + \phi_p B^p) x_t + \varepsilon_t$$

where ε_t is a white-noise variable with mean zero and a constant variance.

[9]The proof is given in BROCKWELL, P.J., and DAVIS, R.A. (2006).

The autoregressive spectrum estimator for the series x_t is defined as: [10]

$$\hat{s}(\omega) = 10 \times \log_{10} \frac{\sigma_x^2}{2\pi |1 - \sum_{k=1}^{p} \hat{\phi}_k e^{-ik\omega}|^2}$$

where:

ω– frequency, $0 \leq \omega \leq \pi$;

σ_x^2 – the innovation variance of the sample residuals;

$\hat{\phi}_k$ – AR(k) coefficient estimates of the linear regression of $x_t - \overline{x}$ on $x_{t-k} - \overline{x}$, $1 \leq k \leq p$.

The autoregressive spectrum estimator is used in the visual spectral analysis tool for detecting significant peaks in the spectrum. The criterion of *visual significance*, implemented in JDemetra+, is based on the range $\hat{s}^{\max} - \hat{s}^{\min}$ of the $\hat{s}(\omega)$ values, where $\hat{s}^{\max} = \max_k \hat{s}(\omega_k)$; $\hat{s}^{\min} = \min_k \hat{s}(\omega_k)$; and $\hat{s}(\omega\_k)$ is k^{th} value of autoregressive spectrum estimator.

The particular value is considered to be visually significant if, at a trading day or at a seasonal frequency ω_k (other than the seasonal frequency $\omega_{60} = \pi$), $\hat{s}(\omega\_k)$ is above the median of the plotted values of $\hat{s}(\omega_k)$ and is larger than both neighbouring values $\hat{s}(\omega_{k-1})$ and $\hat{s}(\omega_{k+1})$ by at least $\frac{6}{52}$ times the range $\hat{s}^{\max} - \hat{s}^{\min}$.

Following the suggestion of SOUKUP, R.J., and FINDLEY, D.F. (1999), JDemetra+ uses an autoregressive model spectral estimator of model order 30. This order yields high resolution of strong components, meaning peaks that are sharply defined in the plot of $\hat{s}(\omega)$ with 61 frequencies. The minimum number of observations needed to compute the spectrum is set to $n = 80$ for monthly data and to $n = 60$ for quarterly series while the maximum number of observations considered for the estimation is 121. Consequently, with these settings it is possible to identify up to 30 peaks in the plot of 61 frequencies. By choosing $\omega_k = \frac{\text{k}}{60}$ for $k = 0, 1, ..., 60$ the density estimates are calculated at exact seasonal frequencies (1, 2, 3, 4, 5 and 6 cycles per year).

The model order can also be selected based on the AIC criterion (in practice it is much lower than 30). A lower order produces the smoother spectrum, but the contrast between the spectral amplitudes at the trading day frequencies and neighbouring frequencies is weaker, and therefore not as suitable for automatic detection.

SOUKUP, R.J., and FINDLEY, D.F. (1999) also explain that the periodogram can be used in the *visual significance* test as it has as good as those of the AR(30) spectrum abilities to detect trading day effect, but also has a greater false alarm rate[11].

comment2: end part theory>spectral analysis>auto-regressive spectrum

[10] Definition from '*X-12-ARIMA Reference Manual*' (2011).

[11] The false alarm rate is defined as the fraction of the 50 replicates for which a visually significant spectral peak occurred at one of the trading day frequencies being considered in the designated output spectra (SOUKUP, R.J., and FINDLEY, D.F. (1999)).

14.2.3 Method 3: Tukey spectrum

14.3 Identification of spectral peaks

comment3: start part theory>spectral analysis>identification of spectral peaks

Identification of seasonal peaks in a Tukey periodogram and in an autoregressive spectrum

In order to decide whether a series has a seasonal component that is predictable (stable) enough, these tests use visual criteria and formal tests for the periodogram. The periodogram is calculated using complete years, so that the set of Fourier frequencies contains exactly all seasonal frequencies[12].

The tests rely on two basic principles:

- The peaks associated with seasonal frequencies should be larger than the median spectrum for all frequencies and;
- The peaks should exceed the spectrum of the two adjacent values by more than a critical value.

  JDemetra+ performs this test on the original series. If these two requirements are met, the test results are displayed in green. The statistical significance of each of the seasonal peaks (i.e. frequencies $\frac{\pi}{6}$, $\frac{\pi}{3}$, $\frac{\pi}{2}$, $\frac{2\pi}{3}$ and $\frac{5\pi}{6}$ corresponding to 1, 2, 3, 4 and 5 cycles per year) is also displayed. The seasonal and trading days frequencies depends on the frequency of time series. They are shown in the table below. The symbol d denotes a default frequency and is described below the table.

The seasonal and trading day frequencies by time series frequency

{: .table .table-style} |**Number of months per full period** | **Seasonal frequency** | **Trading day frequency (radians)**| |————————————| ————————————————————————-| ————————————| |12 | $\frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}, \pi$ | d, 2.714| |6 | $\frac{\pi}{3}, \frac{2\pi}{3}, \pi$ | d |4 | $\frac{\pi}{2}, \pi$ | d, 1.292, 1.850, 2.128| |3 | π | d| |2 | π | d|

The calendar (trading day or working day) effects, related to the variation in the number of different days of the week per period, can induce periodic patterns in the data that can be similar to those resulting from pure seasonal effects. From the theoretical point of view, 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 the year of 365.25 days is equal to $\frac{365.25}{12} = 30.4375 days$). This effect occurs $\frac{365.25}{12} \times \frac{1}{7} = 4.3482 times per month$: one time for each one of the four complete weeks of each month, and a residual of 0.3482 cycles per month, i.e. $0.3482 \times 2\pi = 2.1878 radians$. This turns out to be a fundamental frequency for the effects associated with monthly data. In JDemetra+

[12] For definition of the periodogram and Fourier frequencies see section Spectral Analysis

the fundamental frequency corresponding to 0.3482 cycles per month is used in place of the closest frequency $\frac{\mathrm{k}}{60}$. Thus, the quantity $\frac{\pi\times 42}{60}$ is replaced by $\omega_{42} = 0.3482 \times 2\pi = 2.1878$. The frequencies neighbouring ω_{42}, i.e. ω_{41} and ω_{43} are set to, respectively, $2.1865 - \frac{1}{60}$ and $2.1865 + \frac{1}{60}$.

The default frequencies (d) for calendar effect are: 2.188 (monthly series) and 0.280 (quarterly series). They are computed as:

$$\omega_{\text{ce}} = \frac{2\pi}{7}\left(n - 7 \times \left[\frac{n}{7}\right]\right)$$

, [1]

where:

$n = \frac{365.25}{s}$, $s = 4$ for quarterly series and $s = 12$ for monthly series.

Other frequencies that correspond to trading day frequencies are: 2.714 (monthly series) and 1.292, 1.850, 2.128 (quarterly series).

In particular, the calendar frequency in monthly data (marked in red on the figure below) is very close to the seasonal frequency corresponding to 4 cycles per year $_{40} = \frac{2}{3}\pi = 2.0944$.

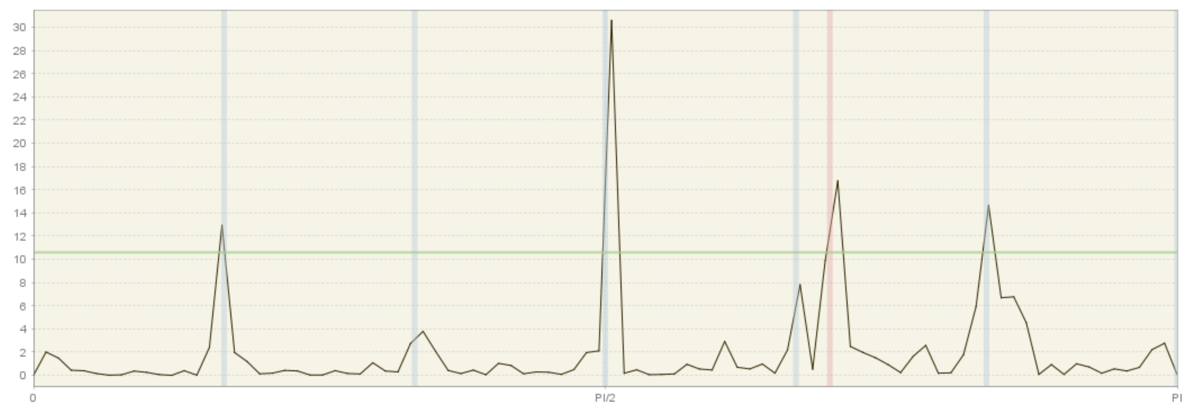

Figure 14.1: Text

Periodogram with seasonal (grey) and calendar (red) frequencies highlighted

This implies that it may be hard to disentangle both effects using the frequency domain techniques.

comment3: end part theory>spectral analysis>identification of spectral peaks

14.3.0.1 in Tukey spectrum

comes from Identification of seasonal peaks in a Tukey spectrum

14.3.0.2 Tukey Spectrum definition

The Tukey spectrum belongs to the class of lag-window estimators. A lag window estimator of the spectral density $f(\omega) = \frac{1}{2\pi}\sum_{k<-\infty}^{\infty}\gamma(k)e^{ik\omega}$ is defined as follows:

$$\hat{f}_L(\omega) = \frac{1}{2\pi}\sum_{|h|\leq r} w(h/r)\hat{\gamma}(h)e^{ih\omega}$$

where $\hat{\gamma}(.)$ is the sample autocovariance function, $w(.)$ is the lag window, and r is the truncation lag. $|w(x)|$ is always less than or equal to one, $w(0) = 1$ and $w(x) = 0$ for $|x| > 1$. The simple idea behind this formula is to down-weight the autocovariance function for high lags where $\hat{\gamma}(h)$ is more unreliable. This estimator requires choosing r as a function of the sample size such that $r/n \to 0$ and $r \to \infty$ when $n \to \infty$. 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 $w(h/r) = 0.5(1 + cos(\pi h/r))$ if $|h/r| \leq 1$ and 0 otherwise.

The choice of large truncation lags r decreases the bias, of course, but it also increases the variance of the spectral estimate and decreases the bandwidth.

JDemetra+ allows the user to modify all the parameters of this estimator, including the window function.

14.3.0.3 Graphical Test

The current JDemetra+ implementation of the seasonality test is based on a $F(d_1, d_2)$ approximation that has been originally 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 r and sample size. Following this approach, we determine visually significant peaks for a frequency ω_j when

$$\frac{2f_x(\omega_j)}{\left[f_x(\omega_{j+1}) + f_x(\omega_{j-1})\right]} \geq CV(\omega_j)$$

where $CV(\omega_j)$ is the critical value of a $F(d_1, d_2)$ distribution, where the degrees of freedom are determined using simulations. For $\omega_j = \pi$, we have a significant peak when $\frac{f_x(\omega_{[n/2]})}{\left[f_x(\omega_{[(n-1)/2]})\right]} \geq CV(\omega_j)$

Two significant levels for this test are considered: $\alpha = 0.05$ (code "t") and $\alpha = 0.01$ (code "T").

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 $r = 79$ for a sample size within the interval $\in [80, 119]$ and $r = 112$ for $n \in [120, 300]$. The F approximation is less accurate for sample sizes larger than 300. For quarterly data, $r = 44$, but there are no recommendations regarding the required sample size.

14.3.0.4 Use

The test can be applied directly to any series by selecting the option *Statistical Methods » Seasonal Adjustment » Tools » Seasonality Tests*. This is an example of how results are displayed for the case of a monthly series:

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

AT.AT.AT.AT.AT.A-

Figure 14.2: tktest

JDemetra+ considers critical values for $\alpha = 1\%$ (code "T") and $\alpha = 5\%$ (code "t") at each one of the seasonal frequencies represented in the table below, e.g. frequencies $\frac{\pi}{6}$, $\frac{\pi}{3}$, $\frac{\pi}{2}$, $\frac{2\pi}{3}$ and $\frac{5\pi}{6}$ corresponding to 1, 2, 3, 4, 5 and 6 cycles per year in this example, since we are dealing with monthly data. The codes "a" and "A" correpond to the so-called AR spectrum, so ignore them for the moment.

The seasonal and trading day frequencies by time series frequency

Number of months per full period	Seasonal frequency	Trading day frequency (radians)
12	$\frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}, \pi$	d, 2.714
6	$\frac{\pi}{3}, \frac{2\pi}{3}, \pi$	d
4	$\frac{\pi}{2}, \pi$	d, 1.292, 1.850, 2.128
3	π	d
2	π	d

Currently, only seasonal frequencies are tested, but the program allows you to manually plot the Tukey spectrum and focus your attention on both seasonal and trading day frequencies.

14.3.0.5 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

14.3.0.6 in AR Spectrum definition

comes from: "Identification of seasonal peaks in autoregressive spectrum"

The estimator of the spectral density at frequency $\lambda \in [0, \pi]$ will be given by the assumption that the series will follow an AR(p) process with large p. The spectral density of such model, with an innovation variance $var(x_t) = \sigma_x^2$, is expressed as follows:

$$10 \times log_{10} f_x(\lambda) = 10 \times log_{10} \frac{\sigma_x^2}{2\pi \left|\phi(e^{i\lambda})\right|^2} = 10 \times log_{10} \frac{\sigma_x^2}{2\pi \left|1 - \sum_{k=1}^{p} \phi_k e^{ik\lambda})\right|^2}$$

where ϕ_k denotes the AR(k) coefficient, and $e^{-ik\lambda} = cos\ (-ik\lambda) + isin\ (-ik\lambda)$.

Soukup and Findely (1999) suggest the use of p=30, 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 *n=80* for monthly data (or *n=60*) for quarterly series. In turn, the maximum number of observations considered for the estimation is *n=121.* This choice offers enough resolution, being able to identify a maximum of 30 peaks in a plot of 61 frequencies: by choosing $\lambda_j = \pi j/60$,for $j = 0, 1, ..., 60$, 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 $2\pi(x/12)$.

The traditional trading day frequency corresponding to 0.348 cycles per month is used in place of the closest frequency $\pi j/60$. Thus, we replace $\pi 42/60$ by $\lambda_{42} = 0.348 \times 2\pi = 2.1865$. The frequencies neighbouring λ_{42} are set to $\lambda_{41} = 2.1865 - 1/60$ and $\lambda_{43} = 2.1865 + 1/60$. The periodogram below illustrates the proximity of this trading day frequency λ_{42} (red shade) and the frequency corresponding to 4 cycles per year $\lambda_{40} = 2.0944$. This proximity is precisely what poses the identification problems: the AR spectrum boils down to a smoothed version of the periodogram and the contribution of the of the trading day frequency may be obscured by the leakage resulting from the potential seasonal peak at λ_{40}, and vice-versa.

Periodogram with seasonal (grey) and calendar (red) frequencies highlighted

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.

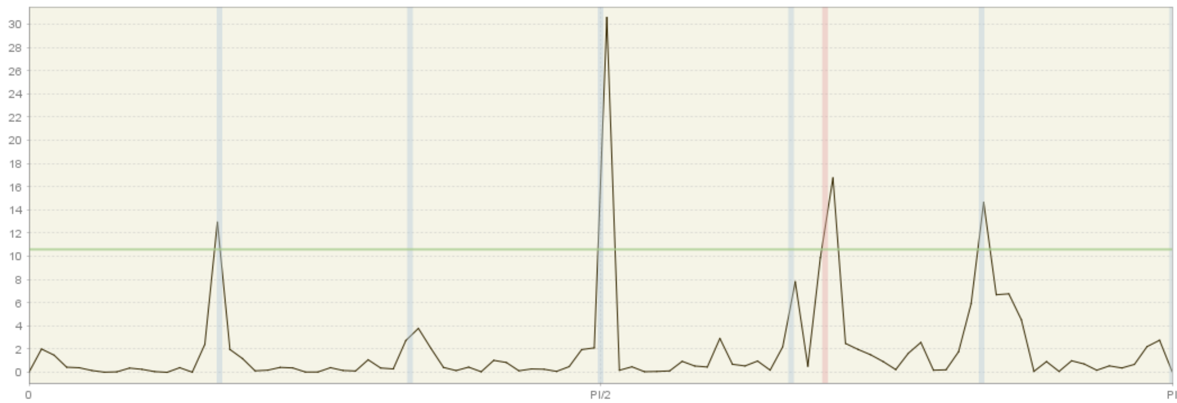

Figure 14.3: Text

14.3.0.7 Graphical Test

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

- $\frac{f_x(\lambda_j)-\max\{f_x(\lambda_{j+1}),f_x(\lambda_{j-1})\}}{[\max_k f_x(\lambda_k)-\min_i f_x(\lambda_i)]} \geq CV(\lambda_j)$, where $CV(\lambda_j)$ can be set equal to $6/52$ for all j
- $f_x(\lambda_j) > median_j \left\{f_x(\lambda_j)\right\}$, which guarantees $f_x(\lambda_j)$ it is not a local peak.

The first condition implies that if we divide the range $\max_k f_x(\lambda_k) - \min_i f_x(\lambda_i)$ in 52 parts (traditionally represented by stars) the height of each pick should be at least 6 stars.

14.3.0.8 Use

The test can be applied directly to any series by selecting the option *Statistical Methods » Seasonal Adjustment » Tools » Seasonality Tests*. This is an example of how results are displayed for the case of a monthly series:

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

AT.AT.AT.AT.AT.A-

Figure 14.4: artest

JDemetra+ considers critical values for $\alpha = 1\%$ (code "A") and $\alpha = 5\%$ (code "a") at each one of the seasonal frequencies represented in the table below, e.g. frequencies $\frac{\pi}{6}$, $\frac{\pi}{3}$, $\frac{\pi}{2}$, $\frac{2\pi}{3}$ and $\frac{5\pi}{6}$

corresponding to 1, 2, 3, 4, 5 and 6 cycles per year in this example, since we are dealing with monthly data. The codes "t" and "T" correpond to the so-called Tukey spectrum, so ignore them for the moment.

The seasonal and trading day frequencies by time series frequency

{: .table .table-style} |**Number of months per full period** | **Seasonal frequency** | **Trading day frequency (radians)**| |————————————| ————————————-| ————————————| |12 | $\frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}, \pi$ | d, 2.714| |6 | $\frac{\pi}{3}, \frac{2\pi}{3}, \pi$ | d |4 | $\frac{\pi}{2}, \pi$ | d, 1.292, 1.850, 2.128| |3 | π | d| |2 | π | d|

Currently, only seasonal frequencies are tested, but the program allows you to manually plot the AR spectrum and focus your attention on both seasonal and trading day frequencies. Agustin Maravall has conducted a simulation experiment to calculate $CV(\lambda_{42})$ (trading day frequency) and proposes to set for all j equal to the critical value associated to the trading frequency, but this is currently not part of the current automatic testing procedure of JDemetra+.

14.3.0.9 References

- 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.

14.3.0.10 in a Periodogram

comes from: Identification of seasonal peaks in periodogram

The periodogram $I(\omega_j)$ of $\mathbf{X} \in \mathbb{C}^n$ is defined as the squared of the Fourier transform

$$I(\omega_j) = a_j^2 = n^{-1} \left| \sum_{t=1}^{n} \mathbf{X_t} e^{-it\omega_j} \right|^2 ,$$

where the Fourier frequencies ω_j are given by multiples of the fundamental frequency $\frac{2\pi}{n}$:

$$\omega_j = \frac{2\pi j}{n}, -\pi < \omega_j \leq \pi$$

An orthonormal basis in $\mathbb{R}^n$:

$$\left\{ e_0, \quad c_1, s_1, \quad \ldots \quad , \quad c_{[(n-1)/2]}, s_{[(n-1)/2]} \quad , \quad e_{n/2} \right\},$$

where $e_{n/2}$ is excluded if n is odd,
can be used to project the data and obtain the spectral decomposition

Thus, the periodogram is given by the projection coefficients and represents the contribution of the jth harmonic to the total sum of squares, as illustrated by Brockwell and Davis (1991):

Source	Degrees of freedom
Frequency ω_0	1
Frequency ω_1	2
$\vdots$	$\vdots$
Frequency ω_k	2
$\vdots$	$\vdots$
Frequency $\omega_{n/2} = \pi$	1
(excluded if n is odd)	
=========	======
Total	n

In JDemetra+, the periodogram of $\mathbf{X} \in \mathbb{R}^n$ is computed for the standardized time series.

14.3.0.11 Defining a F-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 above to derive a test. Thus, under the null hypothesis:

- $2I(\omega_k) = \|P_{\bar{sp}_{\{c_k,s_k\}}}\mathbf{X}\|^2 \sim \sigma^2\chi^2(2)$, for Fourier frequencies $0 < \omega_k = 2\pi k/n < \pi$
- $I(\pi) = \|P_{\bar{sp}_{\left\{e_{n/2}\right\}}}\mathbf{X}\|^2 \sim \sigma^2\chi^2(1)$, for π

Because $I(\omega_k)$ is independent from the projection error sum of squares, we can define our F-test statistic as follows:

- $\frac{2I(\omega_k)}{\|\mathbf{X}-P_{\bar{sp}_{\{e_0,c_k,s_k\}}}\mathbf{X}\|^2}\frac{n-3}{2} \sim F(2, n-3)$, for Fourier frequencies $0 < \omega_k = 2\pi k/n < \pi$
- $\frac{I(\pi)}{\|\mathbf{X}-P_{\bar{sp}_{\left\{e_0,e_{n/2}\right\}}}\mathbf{X}\|^2}\frac{n-2}{1} \sim F(1, n-2)$, for π

where - $\|\mathbf{X} - P_{\bar{sp}_{\{e_0,c_k,s_k\}}}\mathbf{X}\|^2 = \sum_{i=1}^{n} \mathbf{X_i^2} - I(0) - 2I(\omega_k) \sim \sigma^2\chi^2(n-3)$ for Fourier frequencies $0 < \omega_k = 2\pi k/n < \pi$ - $\|\mathbf{X} - P_{\bar{sp}_{\{e_0,e_{n/2}\}}}\mathbf{X}\|^2 = \sum_{i=1}^{n} \mathbf{X_i^2} - I(0) - I(\pi) \sim \sigma^2\chi^2(n-2)$ for π

Thus, we reject the null if our F-test statistic computed at a given seasonal frequency (different from π) is larger than $F_{1-\alpha}(2, n-3)$. If we consider π, our test statistic follows a $F_{1-\alpha}(1, n-2)$ distribution.

14.3.0.12 Seasonality test

The implementation of JDemetra+ considers simultaneously the whole set of seasonal frequencies (1, 2, 3, 4, 5 and 6 cycles per year). Thus, the resulting test-statistic is:

$$\frac{2I(\pi/6) + 2I(\pi/3) + 2I(2\pi/3) + 2I(5\pi/6) + \delta I(\pi)}{\left\|\mathbf{X} - P_{\bar{sp}_{\{e_0,c_1,s_1,c_2,s_2,c_3,s_3,c_4,s_4,c_5,s_5,\delta e_{n/2}\}}}\mathbf{X}\right\|^2}\frac{n-12}{11} \sim F(11-\delta, n-12+\delta)$$

where $\delta = 1$ if n is even and 0 otherwise.

In small samples, the test performs better when the periodogram is evaluated as the exact seasonal frequencies. 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 leakage problem.

Example of how results are displayed:

5. Periodogram

Test on the sum of the values of a periodogram at seasonal frequencies

Seasonality present

Distribution: F with 11 degrees of freedom in the nominator and 180 degrees of freedom in the denominator
Value: 45.1387
PValue: 0.0000

Figure 14.5: periodtest

14.3.0.13 References

Brockwell, P.J., and R.A. Davis (1991). Times Series: Theory and Methods. Springer Series in Statistics.

14.4 Spectral graphs

probably moove this part to GUI (Tools), just leave a link

comment3: start part case studies > spectral graphs

This scenario is designed for advanced users interested in an in-depth analysis of time series in the frequency domain using three spectral graphs. Those graphs can also be used as a complementary analysis for a better understanding of the results obtained with some of the tests described above.

Economic time series are usually presented in a time domain (X-axis). However, for analytical purposes it is convenient to convert the series to a frequency domain due to the fact that any stationary time series can be expressed as a combination of cosine (or sine) functions. These functions are characterized with different periods (amount of time to complete a full cycle) and amplitudes (maximum/minimum value during the cycle).

The tool used for the analysis of a time series in a frequency domain is called a spectrum. The peaks in the spectrum indicate the presence of cyclical movements with periodicity between two months and one year. A seasonal series should have peaks at the seasonal frequencies. Calendar adjusted data are not expected to have peak at with a calendar frequency.

The periodicity of the phenomenon at frequency f is $\frac{2\pi}{f}$. It means that for a monthly time series the seasonal frequencies $\frac{\pi}{6}$, $\frac{\pi}{3}$, $\frac{\pi}{2}$, $\frac{2\pi}{3}$, $\frac{5\pi}{6}$ and π correspond to 1, 2, 3, 4, 5 and 6 cycles per year. For example, the frequency $\frac{\pi}{3}$ corresponds to a periodicity of 6 months (2 cycles per year are completed). For the quarterly series there are two seasonal frequencies: $\frac{\pi}{2}$ (one cycle per year) and π (two cycles per year). A peak at the zero frequency always corresponds to the trend component of the series. Seasonal frequencies are marked as grey vertical lines, while violet vertical lines represent the trading-days frequencies. The trading day frequency is 0.348 and derives from the fact that a daily component which repeats every seven days goes through 4.348 cycles in a month of average length 30.4375 days. It is therefore seen to advance 0.348 cycles per month when the data are obtained at twelve equally spaced times in 365.25 days (the average length of a year).

The interpretation of the spectral graph is rather straightforward. When the values of a spectral graph for low frequencies (i.e. one year and more) are large in relation to its other values it means that the long-term movements dominate in the series. When the values of a spectral graph for high frequencies (i.e. below one year) are large in relation to its other values it means that the series are rather trendless and contains a lot of noise. When the values of a spectral graph are distributed randomly around a constant without any visible peaks, then it is highly probable that the series is a random process. The presence of seasonality in a time series is manifested in a spectral graph by the peaks on the seasonal frequencies.

Spectral graphs in GUI

Auto-regressive spectrum's properties

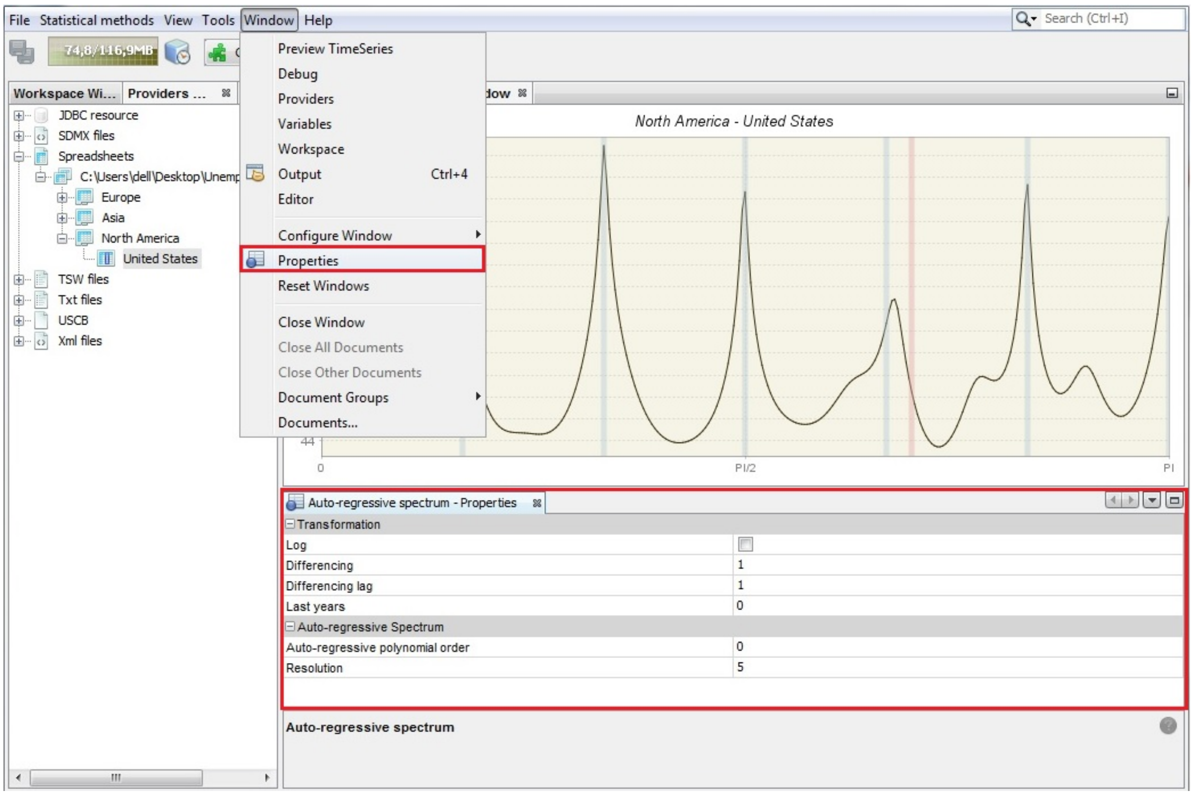

Figure 14.6: Text

1. The spectral graphs are available from: *Tools* → *Spectral analysis*.

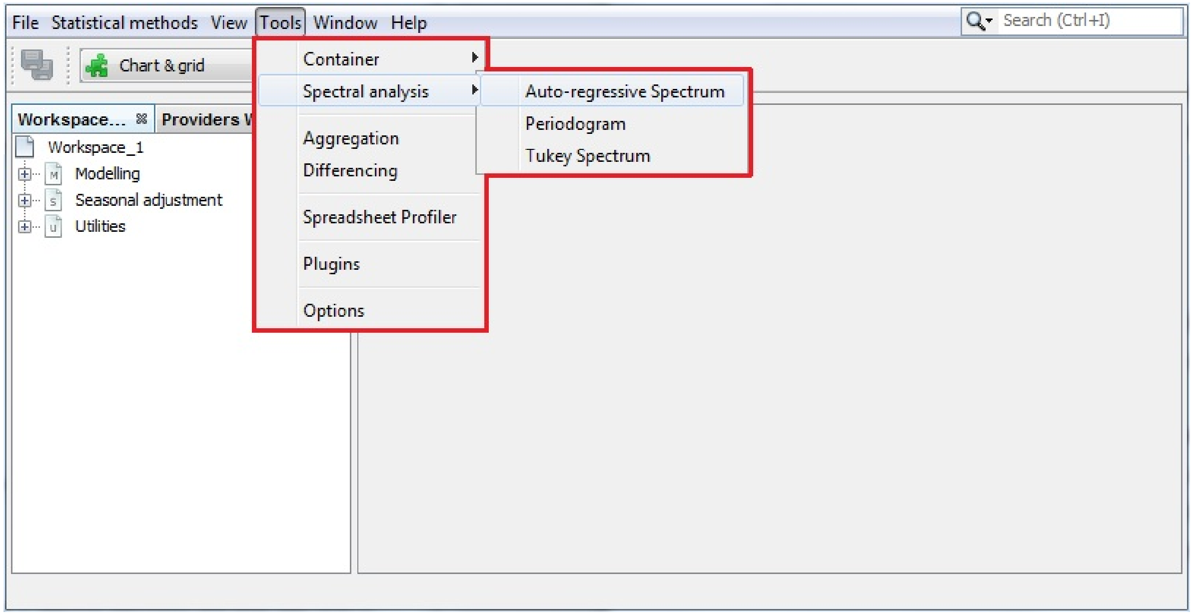

Figure 14.7: Text

Tools for spectral analysis

2. When the first option is chosen JDemetra+ displays an empty *Auto-regressive spectrum* window. To start an analysis drag a single time series from the *Providers* window and drop it into the *Drop data here* area.

Launching an auto-regressive spectrum

3. An auto-regressive spectrum graph available in JDemetra+ is based on the relevant tool from the X-13ARIMA-SEATS program. It shows the spectral density (spectrum) function, which reformulates the content of the stationary time series' autocovariances in terms of amplitudes at frequencies of half a cycle per month or less. The number of observations, data transformations and other options such as the specification of the frequency grid and the order of the autoregressive polynomial (30 by default) can be specified by opening the *Window* → *Properties* from the main menu.

The *Auto-regressive - Properties* window contains the following options:

- **Log** - a log transformation of a time series;
- **Differencing** - transforms a data by calculating a regular (order 1,2..) or seasonal (order 4, 12, depending on the time series frequency) differences;

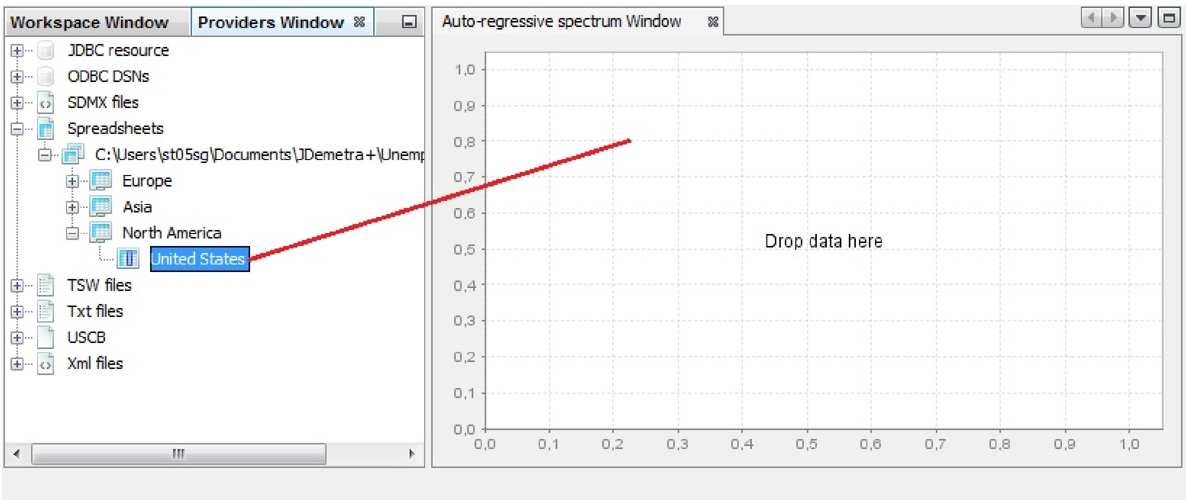

Figure 14.8: Text

- **Differencing lag** - the number of lags that the program will use to take differences. For example, if *Differencing lag = 3* then the differencing filter does not apply to the first lag (default) but to the third lag.
- **Last years** - a number of years at the end of the time series taken to produce autoregresive spectrum. By default, it is 0, which means that the whole time series is considered.
- **Auto-regressive polynomial order** - the number of lags in the AR model that is used to estimate the spectral density. By default, the order of the autoregressive polynomial is set to 30 lags.
- **Resolution** - the value 1 plots the spectral density estimate for the frequencies $\omega_j = \frac{2\pi j}{n}$, where $n \in (-\pi; \pi)$ is the size of the sample used to estimate the AR model. Increasing this value, which is set to 5 by default, will increase the precision of this grid.

4. The seasonality test described above uses an empirical criterion to check whether the series has a seasonal component that is predictable (stable) enough that it can be estimated with reasonable success. The peak in the auto-regressive spectrum has to be greater than the median of the 61 spectrum ordinates and has to exceed the two adjacent spectral values by more than a critical value. When such a case is detected, the test results are displayed in green.

   An example of an-auto-regressive spectrum

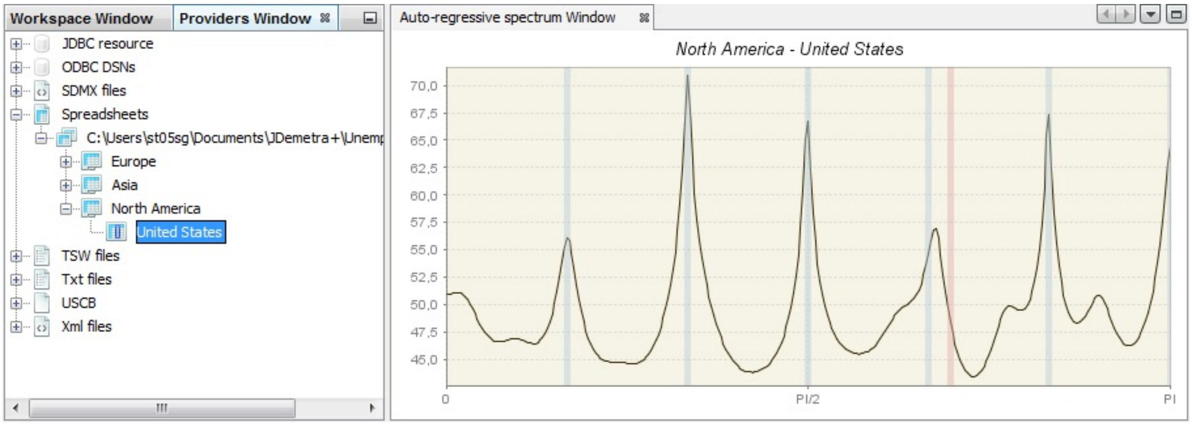

Figure 14.9: Text

5. The second spectral graph is a periodogram. To perform the analysis of a single time series using this tool, choose *Tools* →*Spectral analysis* → *Periodogram* and drag and drop a series from the *Providers* window to the empty *Periodogram* window.

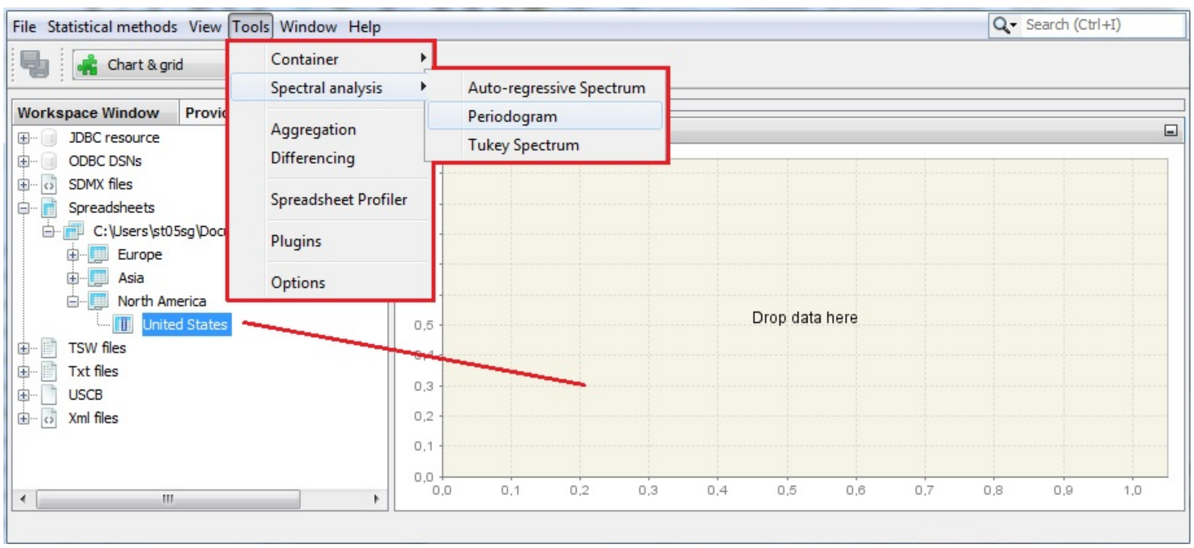

Figure 14.10: Text

Launching a periodogram

6. The sample size and data transformations can be specified by opening the *Window* → *Properties*, in the main menu. The *Periodogram - Properties* window contains the following options:

   - **Log** - a log transformation of a time series;

- **Differencing** - transforms the data by calculating regular (order 1,2..) or seasonal (order 4, 12, depending on the time series frequency) differences;
- **Differencing lag** - the number of lags that you will use to take differences. For example, if *Differencing lag = 3* then the differencing filter does not apply to the first lag (default) but to the third lag.
- **Last years** - the number of years at the end of the time series taken to produce periodogram. By default it is 0, which means that the whole time series is considered.

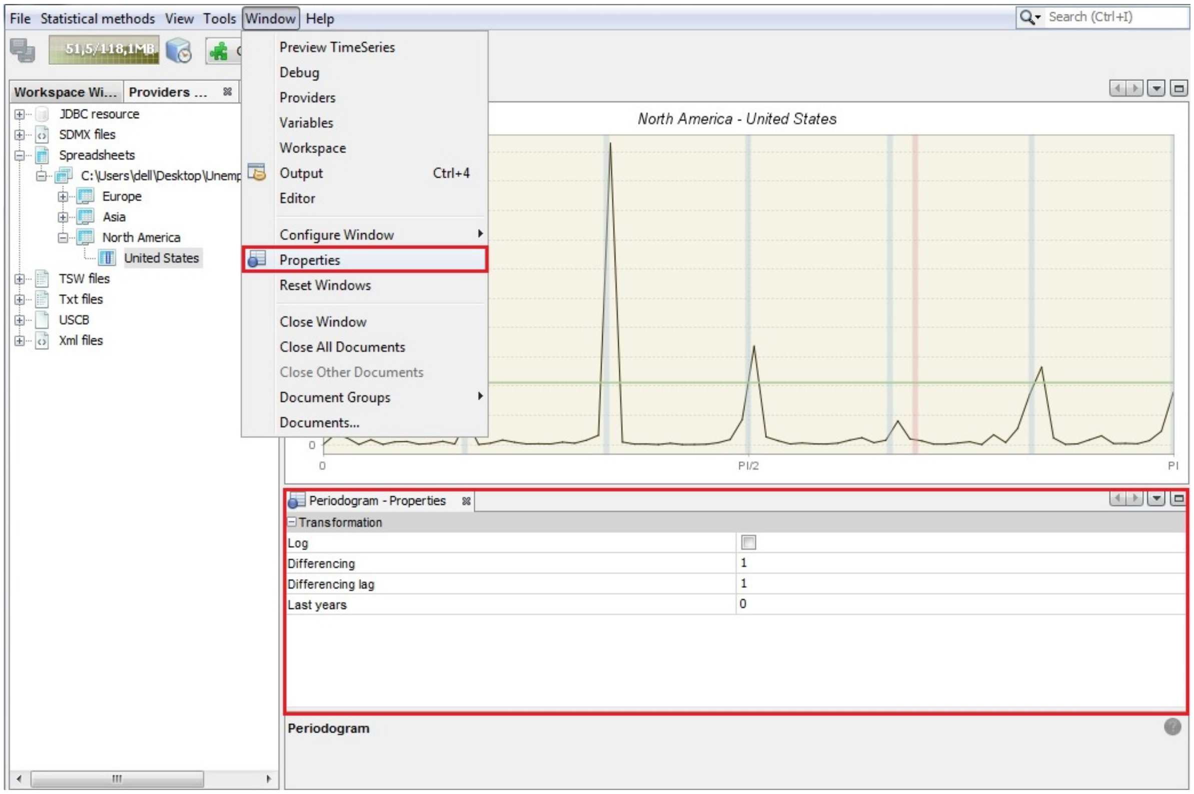

Figure 14.11: Text

Periodogram's properties

7. The periodogram was one of the earliest tools used for the analysis of time series in the frequency domain. It enables the user to identify the dominant periods (or frequencies) of a time series. In general, the periodogram is a wildly fluctuating estimate of the spectrum with a high variance and is less stable than an auto-regressive spectrum.

An example of a periodogram

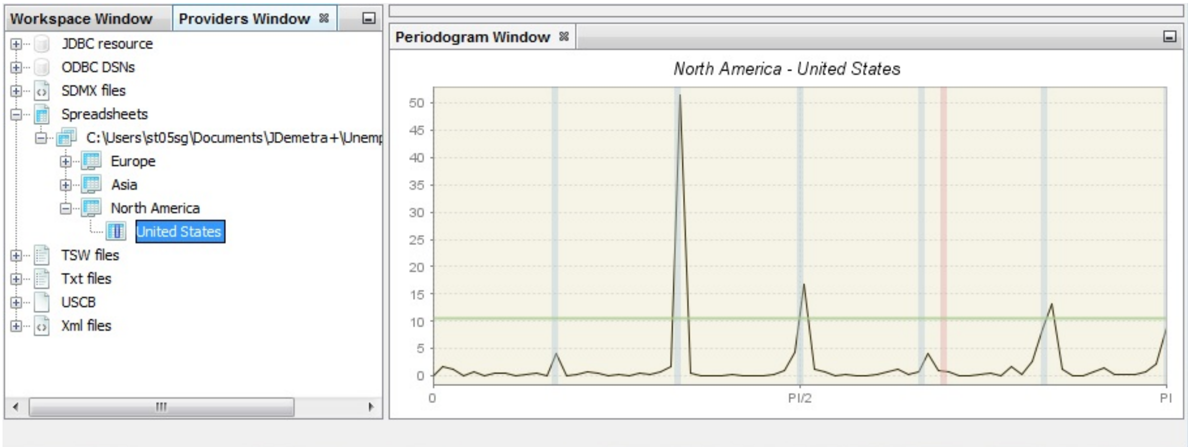

Figure 14.12: Text

8. The third spectral graph is the Tukey spectrum. To perform the analysis of time series using this tool, choose *Tools* → *Spectral analysis* → *Tukey spectrum* and drag and drop a single series from the *Providers* window to the empty *Periodogram* window.

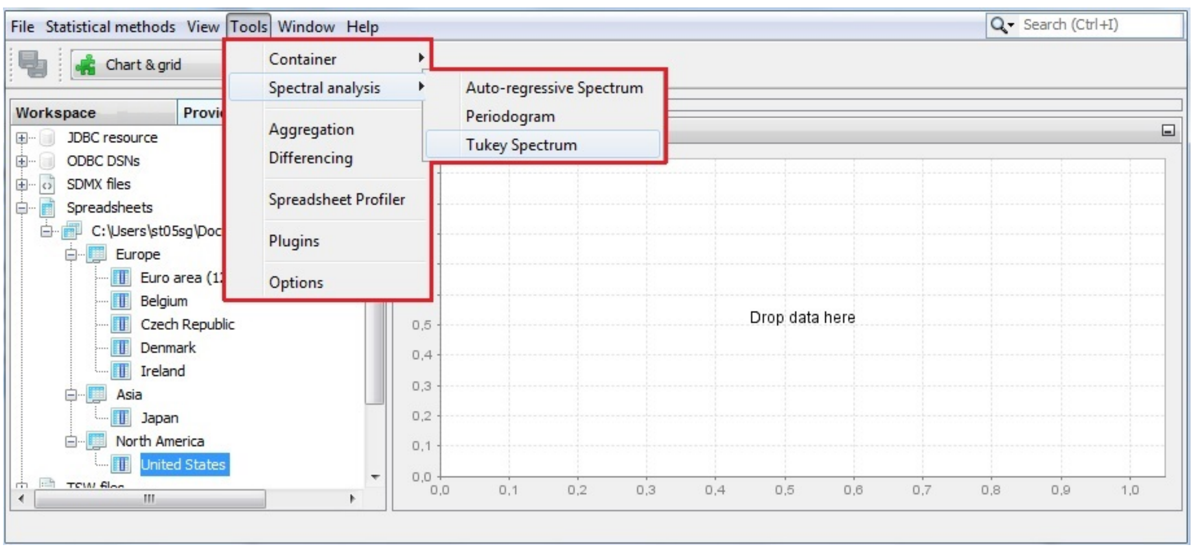

Figure 14.13: Text

Launching a Tukey spectrum

9. The Tukey spectrum estimates the spectral density by smoothing the periodogram.

An example of a Tukey spectrum

10. The options for the Tuckey window can be specified by opening the *Window* → *Properties*

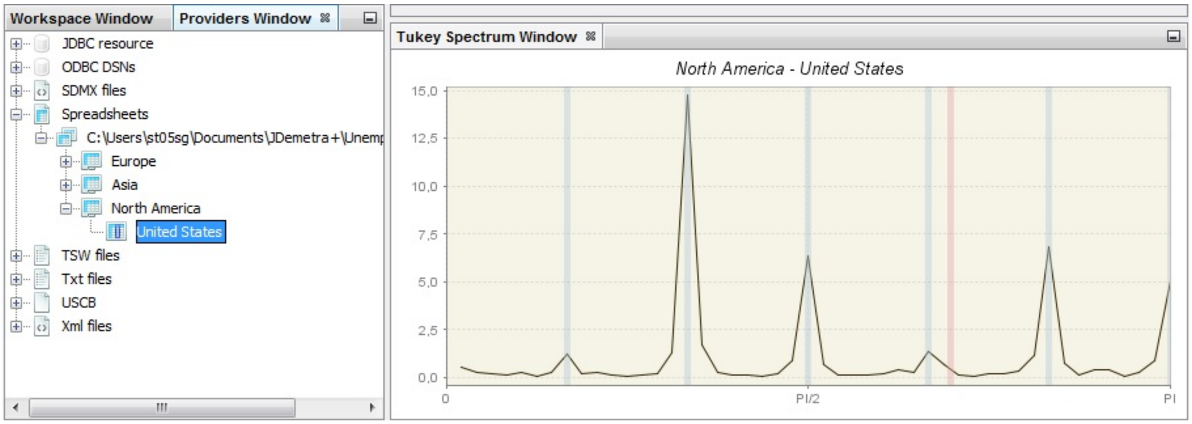

Figure 14.14: Text

from the main menu. The *Periodogram - Properties* window contains the following options:

- **Log** - a log transformation of a time series.
- **Differencing** - transforms the data by calculating regular (order 1, 2..) or seasonal (order 4, 12, depending on the time series frequency) differences.
- **Differencing lag** - the number of lags that you will use to take differences. For example, if *Differencing lag = 3* then the differencing filter does not apply to the first lag (default) but to the third lag.
- **Taper part** – parameter larger than 0 and smaller or equal to one that shapes the curvature of the smoothing function that is applied to the auto-covariance function.
- **Window length** – the size of the window that is used to smooth the auto-covariance function. A value of zero includes the whole series.
- **Window type** – it refers to the weighting scheme that it is used to smooth the auto-covariance function. The available windows types (*Square*, *Welch*, *Tukey*, *Barlett*, *Hamming*, *Parzen*) are suitable to estimate the spectral density.

Tukey spectrum's properties

comment3: end part case studies > spectral graphs

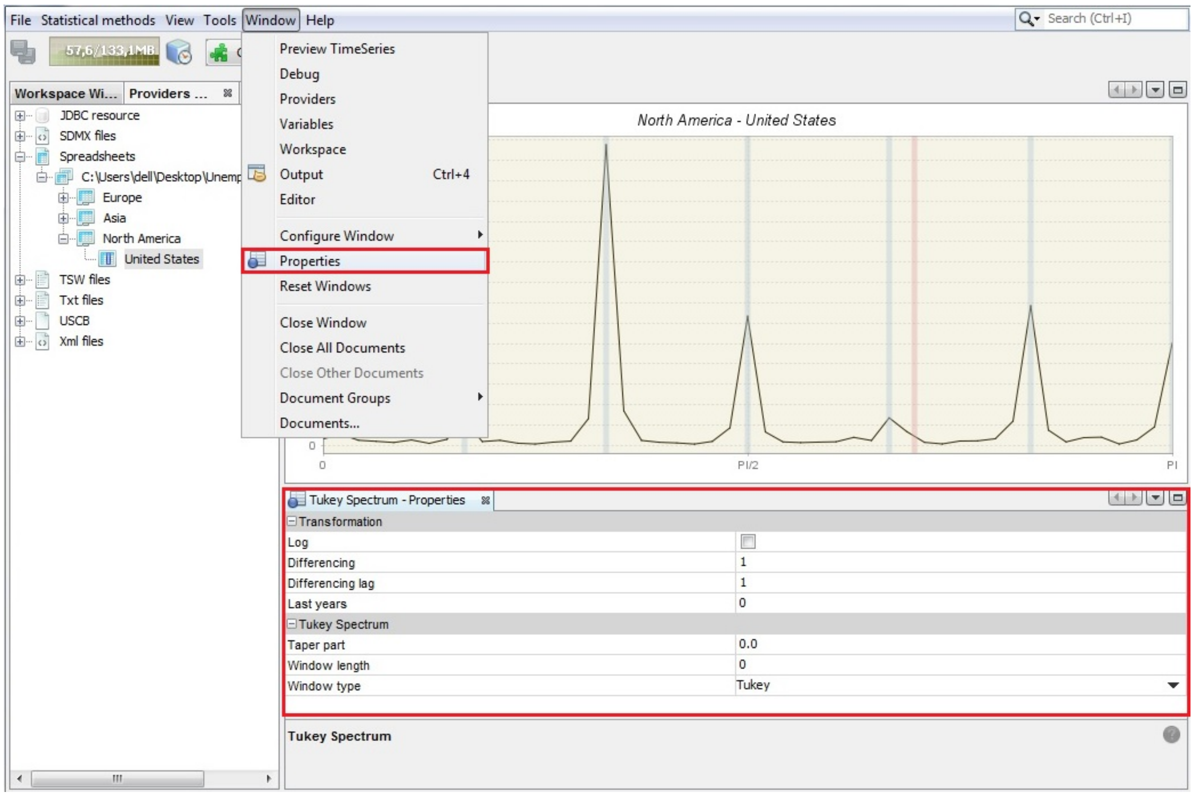

Figure 14.15: Text

15 Reg-Arima models

15.1 Overview

The primary aim of seasonal adjustment is to remove the unobservable seasonal component from the observed series. The decomposition routines implemented in the seasonal adjustment methods make specific assumptions concerning the input series. One of the crucial assumptions is that the input series is stochastic, i.e. it is clean of deterministic effects. Another important limitation derives from the symmetric linear filter used in TRAMO-SEATS and X-13ARIMA-SEATS. A symmetric linear filter cannot be applied to the first and last observations with the same set of weights as for the central observations[^1]. Therefore, for the most recent observations these filters provide estimates that are subject to revisions.

To overcome these constrains both seasonal adjustment methods discussed here include a modelling step that aims to analyse the time series development and provide a better input for decomposition purposes. The tool that is frequently used for this purpose is the ARIMA model, as discussed by BOX, G.E.P., and JENKINS, G.M. (1970). However, time series are often affected by the outliers, other deterministic effects and missing observations. The presence of these effects is not in line with the ARIMA model assumptions. The presence of outliers and other deterministic effects impede the identification of an optimal ARIMA model due to the important bias in the estimation of parameters of sample autocorrelation functions (both global and partial)[^3]. Therefore, the original series need to be corrected for any deterministic effects and missing observations. This process is called linearisation and results in the stochastic series that can be modelled by ARIMA.

For this purpose both TRAMO and RegARIMA use regression models with ARIMA errors. With these models TRAMO and RegARIMA also produce forecasts.

16 X-11 decomposition

16.1 Introduction

A complete documentation of the X-11 method is available in LADIRAY, D., and QUENNEVILLE, B. (2001). The X-11 program is the result of a long tradition of non-parametric smoothing based on moving averages, which are weighted averages of a moving span of a time series (see hereafter). Moving averages have two important drawbacks:

- They are not resistant and might be deeply impacted by outliers;
- The smoothing of the ends of the series cannot be done except with asymmetric moving averages which introduce phase-shifts and delays in the detection of turning points.

These drawbacks adversely affect the X-11 output and stimulate the development of this method. To overcome these flaws first the series are modelled with a RegARIMA model that calculates forecasts and estimates the regression effects. Therefore, the seasonal adjustment process is divided into two parts.

- In a first step, the RegARIMA model is used to clean the series from non-linearities, mainly outliers and calendar effects. A global ARIMA model is adjusted to the series in order to compute the forecasts.
- In a second step, an enhanced version of the X-11 algorithm is used to compute the trend, the seasonal component and the irregular component.

The flow diagram for seasonal adjustment with X-13ARIMA-SEATS using the X-11 algorithm.

16.1.0.1 Moving averages

The moving average of coefficient θ_i is defined as:

$$M\left(X_t\right) = \sum_{k=-p}^{+f} \theta_k X_{t+k}$$

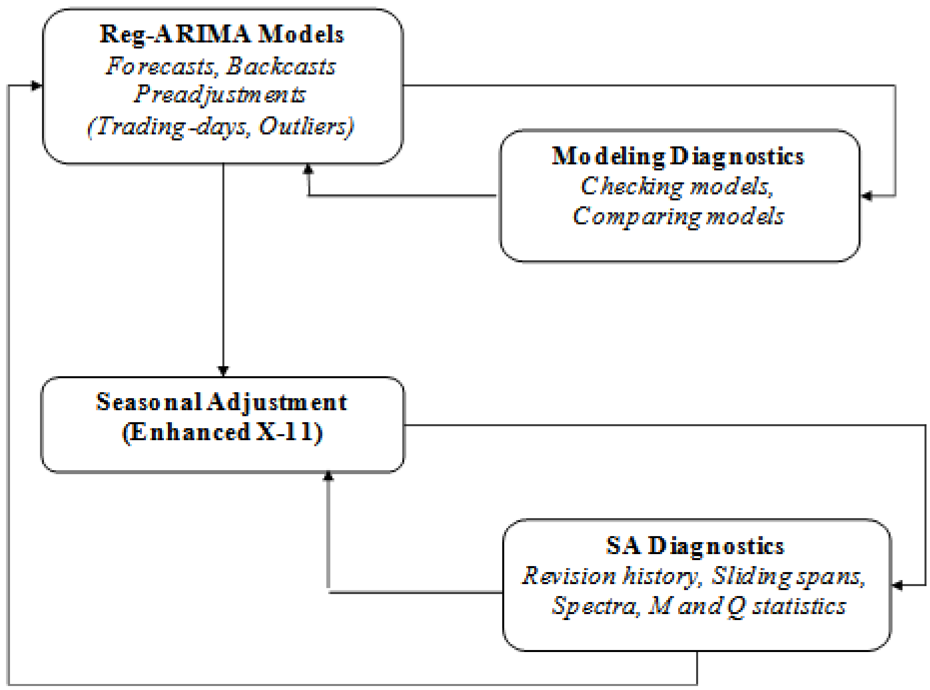

Figure 16.1: Text

The value at time t of the series is therefore replaced by a weighted average of p "past" values of the series, the current value, and f "future" values of the series. The quantity $p + f + 1$ is called the moving average order. When p is equal to f, that is, when the number of points in the past is the same as the number of points in the future, the moving average is said to be centred. If, in addition, $\theta_{-k} = \theta_k$ for any k, the moving average M is said to be symmetric. One of the simplest moving averages is the symmetric moving average of order $P = 2p + 1$ where all the weights are equal to $\frac{1}{P}$.

This moving average formula works well for all time series observations, except for the first p values and last f values. Generally, with a moving average of order $p + f + 1$ calculated for instant t nwith points p in the past and points f in the future, it will be impossible to smooth out the first p values and the last f values of the series because of lack of input to the moving average formula.

In the X-11 method, symmetric moving averages play an important role as they do not introduce any phase-shift in the smoothed series. But, to avoid losing information at the series ends, they are either supplemented by *ad hoc* asymmetric moving averages or applied on the series extended by forecasts.

For the estimation of the seasonal component, X-13ARIMA-SEATS uses $P \times Q$ composite moving averages, obtained by composing a simple moving average of order P, which coefficients are all equal to $\frac{1}{P}$, and a simple moving average of order Q, which coefficients are all equal to $\frac{1}{Q}$.

The composite moving averages are widely used by the X-11 method. For an initial estimation of trend X-11 method uses a 2×4 moving average in case of a quarterly time series while for a monthly time series a 2×12 moving average is applied. The 2×4 moving average is an average of order 5 with coefficients $\frac{1}{8}\{1, 2, 2, 2, 1\}$. It eliminates frequency $\frac{\pi}{2}$ corresponding to period 4 and therefore it is suitable for seasonal adjustment of the quarterly series with a constant seasonality. The 2×12 moving average, with coefficients $\frac{1}{24}\{1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1\}$ that retains linear trends, eliminates order-12 constant seasonality and minimises the variance of the irregular component. The 2×4 and 2×12 moving averages are also used in the X-11 method to normalise the seasonal factors. The composite moving averages are also used to extract the seasonal component. These, which are used in the purely automatic run of the X-11 method (without any intervention from the user) are 3×3, 3×5 and 3×9.

In the estimation of the trend also Henderson moving averages are used. These filters have been chosen for their smoothing properties. The coefficients of a Henderson moving average of order $2p + 1$ may be calculated using the formula:

$$\theta_i = \frac{315\left[(n-1)^2 - i^2\right]\left[n^2 - i^2\right]\left[(n+1)^2 - i^2\right]\left[3n^2 - 16 - 11i^2\right]}{8n\left(n^2-1\right)\left(4n^2-1\right)\left(4n^2-9\right)\left(4n^2-25\right)}$$

,

2

where: $n = p + 2n = p + 2$.

16.1.0.2 The basic algorithm of the X-11 method

The X-11 method is based on an iterative principle of estimation of the different components using appropriate moving averages at each step of the algorithm. The successive results are saved in tables. The list of the X-11 tables displayed in JDemetra+ is included at the end of this section.

The basic algorithm of the X-11 method will be presented for a monthly time series X_t that is assumed to be decomposable into trend, seasonality and irregular component according to an additive model $X_t = TC_t + S_t + I_t$.

A simple seasonal adjustment algorithm can be thought of in eight steps. The steps presented below are designed for the monthly time series. In the algorithm that is run for the quarterly time series the 2×4 moving average instead of the 2×12 moving average is used.

Step 1: Estimation of Trend by $\mathbf{2 \times 12}$ moving average:

$$TC_t^{(1)} = M_{2\times12}(X_t)$$

3

Step 2: Estimation of the Seasonal-Irregular component:

$$(S_t + I_t)^{(1)} = X_t - \mathrm{TC}_t^{(1)}$$

4

Step 3: Estimation of the Seasonal component by

$$\mathbf{3 \times 3}$$

moving average over each month:

$$S_t^{(1)} - M_{3\times3}\left[(S_t + I_t)^{(1)}\right]$$

5

The moving average used here is a 3×3 moving average over 5 terms, with coefficients $\frac{1}{9}\{1,2,3,2,1\}$. The seasonal component is then centred using a 2×12 moving average.

$$\widetilde{S}_t^{(1)} = S_t^{(1)} - M_{2\times12}\left(S_t^{(1)}\right)$$

6

Step 4: Estimation of the seasonally adjusted series:

$$SA_t^{(1)} = (\mathrm{TC}_t + I_t)^{(1)} = X_t - \widetilde{S}_t^{(1)}$$

7

This first estimation of the seasonally adjusted series must, by construction, contain less seasonality. The X-11 method again executes the algorithm presented above, changing the moving averages to take this property into account.

Step 5: Estimation of Trend by 13-term Henderson moving average:

$$TC_t^{(2)} = H_{13}\left(\mathrm{SA}_t^{(1)}\right)$$

8

Henderson moving averages, while they do not have special properties in terms of eliminating seasonality (limited or none at this stage), have a very good smoothing power and retain a local polynomial trend of degree 2 and preserve a local polynomial trend of degree 3.

Step 6: Estimation of the Seasonal-Irregular component:

$$(S_t + I_t)^{(2)} = X_t - \mathrm{TC}_t^{(2)}$$

9

Step 7: Estimation of the Seasonal component by $\mathbf{3 \times 5}$ moving average over each month:

$$S_t^{(2)} - M_{3\times 3}\left[(S_t + I_t)^{(2)}\right]$$

10

The moving average used here is a 3×5 moving average over 7 terms, of coefficients

$$\frac{1}{15}\{1,\ 2,\ 3,\ 3,\ 3,\ 2,\ 1\}$$

and retains linear trends. The coefficients are then normalised such that their sum over the whole 12-month period is approximately cancelled out:

$$\widetilde{S}_t^{(2)} = S_t^{(2)} - M_{2\times 12}\left(S_t^{(2)}\right)$$

11

Step 8: Estimation of the seasonally adjusted series:

$$SA_t^{(2)} = (TC_t + I_t)^{(2)} = X_t - \widetilde{S}_t^{(2)}$$

12

The whole difficulty lies, then, in the choice of the moving averages used for the estimation of the trend in steps 1 and 5 on the one hand, and for the estimation of the seasonal component in steps 3 and 5. The course of the algorithm in the form that is implemented in JDemetra+ is presented in the figure below. The adjustment for trading day effects, which is present in the original X-11 program, is omitted here, as since calendar correction is performed by the RegARIMA model, JDemetra+ does not perform further adjustment for these effects in the decomposition step.

A workflow diagram for the X-11 algorithm based upon training material from the Deutsche Bundesbank

16.1.0.2.1 The iterative principle of X-11

To evaluate the different components of a series, while taking into account the possible presence of extreme observations, X-11 will proceed iteratively: estimation of components, search for disruptive effects in the irregular component, estimation of components over a corrected series, search for disruptive effects in the irregular component, and so on.

The Census X-11 program presents four processing stages (A, B, C, and D), plus 3 stages, E, F, and G, that propose statistics and charts and are not part of the decomposition per se. In stages B, C and D the basic algorithm is used as is indicated in the figure below.

A workflow diagram for the X-11 algorithm implemented in JDemetra+. Source: Based upon training material from the Deutsche Bundesbank

- **Part A: Pre-adjustments**

This part, which is not obligatory, corresponds in X-13ARIMA-SEATS to the first cleaning of the series done using the RegARIMA facilities: detection and estimation of outliers and calendar effects (trading day and Easter), forecasts and backcasts[^61] of the series. Based on these results, the program calculates prior adjustment factors that are applied to the raw series. The series thus corrected, Table B1 of the printouts, then proceeds to part B.

- **Part B: First automatic correction of the series**

This stage consists of a first estimation and down-weighting of the extreme observations and, if requested, a first estimation of the calendar effects. This stage is performed by applying the basic algorithm detailed earlier. These operations lead to Table B20, adjustment values for extreme observations, used to correct the unadjusted series and result in the series from Table C1.

- **Part C: Second automatic correction of the series**

Applying the basic algorithm once again, this part leads to a more precise estimation of replacement values of the extreme observations (Table C20). The series, finally "cleaned up", is shown in Table D1 of the printouts.

- **Part D: Seasonal adjustment**

This part, at which our basic algorithm is applied for the last time, is that of the seasonal adjustment, as it leads to final estimates:

- of the seasonal component (Table D10);
- of the seasonally adjusted series (Table D11);
- of the trend component (Table D12);
- of the irregular component (Table D13).

- **Part E: Components modified for large extreme values**

Parts E includes:

- Components modified for large extreme values;
- Comparison the annual totals of the raw time series and seasonally adjusted time series;
- Changes in the final seasonally adjusted series;
- Changes in the final trend;
- Robust estimation of the final seasonally adjusted series.

The results from part E are used in part F to calculate the quality measures.

- **Part F: Seasonal adjustment quality measures**

Part F contains statistics for judging the quality of the seasonal adjustment. JDemetra+ presents selected output for part F, i.e.:

- M and Q statistics;
- Tables.

- **Part G: Graphics**

Part G presents spectra estimated for:

- Raw time series adjusted a priori (Table B1);
- Seasonally adjusted time series modified for large extreme values (Table E2);
- Final irregular component adjusted for large extreme values (Table E3).

Originally, graphics were displayed in character mode. In JDemetra+, these graphics are replaced favourably by the usual graphics software.

The Henderson moving average and the trend estimation

In iteration B (Table B7), iteration C (Table C7) and iteration D (Table D7 and Table D12) the trend component is extracted from an estimate of the seasonally adjusted series using Henderson moving averages. The length of the Henderson filter is chosen automatically by X-13ARIMA-SEATS in a two-step procedure.

It is possible to specify the length of the Henderson moving average to be used. X-13ARIMA-SEATS provides an automatic choice between a 9-term, a 13-term or a 23-term moving average. The automatic choice of the order of the moving average is based on the value of an indicator called $\frac{\bar{I}}{\bar{C}}$ ratio which compares the magnitude of period-on-period movements in the irregular component with those in the trend. The larger the ratio, the higher the order of the moving average selected. Moreover, X-13ARIMA-SEATS allows the user to choose manually any odd-numbered Henderson moving average. The procedure used in each part is very similar; the only differences are the number of options available and the treatment of the observations in the both ends of the series. The procedure below is applied for a monthly time series.

In order to calculate $\frac{\overline{I}}{\overline{C}}$ ratio a first decomposition of the SA series (seasonally adjusted) is computed using a 13-term Henderson moving average.

For both the trend (C) and irregular (I) components, the average of the absolute values for monthly growth rates (multiplicative model) or for monthly growth (additive model) are computed. They are denoted as $\overline{C}$ and $\overline{I}$, receptively, where $\overline{C} = \frac{1}{n-1}\sum_{t=2}^{n}|C_t - C_{t-1}|$ and $\overline{I} = \frac{1}{n-1}\sum_{t=2}^{n}|I_t - I_{t-1}|$.

Then the value of $\frac{\overline{I}}{\overline{C}}$ ratio is checked and in iteration B:

- If the ratio is smaller than 1, a 9-term Henderson moving average is selected;
- Otherwise, a 13-term Henderson moving average is selected.

Then the trend is computed by applying the selected Henderson filter to the seasonally adjusted series from Table B6. The observations at the beginning and at the end of the time series that cannot be computed by means of symmetric Henderson filters are estimated by ad hoc asymmetric moving averages.

In iterations C and D:

- If the ratio is smaller than 1, a 9-term Henderson moving average is selected;
- If the ratio is greater than 3.5, a 23-term Henderson moving average is selected.
- Otherwise, a 13-term Henderson moving average is selected.

The trend is computed by applying selected Henderson filter to the seasonally adjusted series from Table C6, Table D7 or Table D12, accordingly. At the both ends of the series, where a central Henderson filter cannot be applied, the asymmetric ends weights for the 7 term Henderson filter are used.

16.1.0.2.2 Choosing the composite moving averages when estimating the seasonal component

In iteration D, Table D10 shows an estimate of the seasonal factors implemented on the basis of the modified SI (Seasonal – Irregular) factors estimated in Tables D4 and D9bis. This component will have to be smoothed to estimate the seasonal component; depending on the importance of the irregular in the SI component, we will have to use moving averages of varying length as in the estimate of the Trend/Cycle where the $\frac{\overline{I}}{\overline{C}}$ ratio was used to select the length of the Henderson moving average. The estimation includes several steps.

Step 1: Estimating the irregular and seasonal components

An estimate of the seasonal component is obtained by smoothing, month by month and therefore column by column, Table D9bis using a simple 7-term moving average, i.e. of coefficients $\frac{1}{7}\{1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1\}$. In order not to lose three points at the beginning and end of each column, all columns are completed as follows. Let us assume that the column that corresponds to the month is composed of N values $\{x_1,\ x_2,\ x_3,\ ...x_{N-1},\ x_N\}$. It will be transformed into a series $\left\{x_{-2}, x_{-1}, x_0, x_1,\ x_2,\ x_3,\ ...x_{N-1},\ x_N, x_{N+1},\ x_{N+1},\ x_{N+2},\ x_{N+3}\right\}$ with $x_{-2} = x_{-1} = x_0 = \frac{x_1+x_2+x_3}{3}$ and $x_{N+1} = x_{N+2} = x_{N+3} = \frac{x_N+x_{N-1}+x_{N-2}}{3}$. We then have the required estimates: $S = M_7(D9bis)$ and $I = D9bis - S$.

Step 2: Calculating the Moving Seasonality Ratios

For each i^{th} month the mean annual changes for each component is obtained by calculating

$$\overline{S}_i = \frac{1}{N_i - 1} \sum_{t=2}^{N_i} \left|S_{i,t} - S_{i,t-1}\right|$$

and

$$\overline{I}_i = \frac{1}{N_i - 1} \sum_{t=2}^{N_i} \left|I_{i,t} - I_{i,t-1}\right|$$

, where N_i refers to the number of months iin the data, and the moving seasonality ratio of month i:

$$MSR_i = \frac{\overline{I}_i}{\overline{S}_i}$$

. These ratios are presented in *Details* of the *Quality Measures* node under the *Decomposition (X11)* section. These ratios are used to compare the year-on-year changes in the irregular component with those in the seasonal component. The idea is to obtain, for each month, an indicator capable of selecting the appropriate moving average for the removal of any noise and providing a good estimate of the seasonal factor. The higher the ratio, the more erratic the series, and the greater the order of the moving average should be used. As for the rest, by default the program selects the same moving average for each month, but the user can select different moving averages for each month.

Step 3: Calculating the overall Moving Seasonality Ratio

The overall Moving Seasonality Ratio is calculated as follows:

$$\text{MSR}_i = \frac{\sum_i N_i\ \overline{I}_i}{\sum_i N_i \overline{S}_i}$$

Step 4: Selecting a moving average and estimating the seasonal component

Depending on the value of the ratio, the program automatically selects a moving average that is applied, column by column (i.e. month by month) to the Seasonal/Irregular component in Table D8 modified, for extreme values, using values in Table D9.

The default selection procedure of a moving average is based on the Moving Seasonality Ratio in the following way:

- If this ratio occurs within zone A (MSR < 2.5), a 3×3 moving average is used; if it occurs within zone C ($3.5 <$ MSR < 5.5), a 3×5 moving average is selected; if it occurs within zone E (MSR 6.5), a 3×9 moving average is used;
- If the MSR occurs within zone B or D, one year of observations is removed from the end of the series, and the MSR is recalculated. If the ratio again occurs within zones B or D, we start over again, removing a maximum of five years of observations. If this does not work, i.e. if we are again within zones B or D, a 3×5 moving average is selected.

The chosen symmetric moving average corresponds, as the case may be 5 (3×3), 7 (3×5) or 11 (3×9 3×9) terms, and therefore does not provide an estimate for the values of seasonal factors in the first 2 (or 3 or 5) and the last 2 (or 3 or 5) years. These are then calculated using associated asymmetric moving averages.

Moving average selection procedure, source: DAGUM, E. B.(1999)

16.1.0.2.3 Identification and replacement of extreme values

X-13ARIMA-SEATS detects and removes outliers in the RegARIMA part. However, if there is a seasonal heteroscedasticity in a time series i.e. the variance of the irregular component is different in different calendar months. Examples for this effect could be the weather and snow-dependent output of the construction sector in Germany during winter, or changes in Christmas allowances in Germany and resulting from this a transformation in retail trade turnover before Christmas. The ARIMA model is not on its own able to cope with this characteristic. The practical consequence is given by the detection of additional extreme values by X-11. This may not be appropriate if the seasonal heteroscedasticity is produced by political interventions or other influences. The ARIMA models assume a constant variance and are therefore not by themselves able to cope with this problem. Choosing longer (in the case of diverging weather conditions in the winter time for the construction sector) or shorter filters (in the case of a changing pattern of retail trade turnover in the Christmas time) may be reasonable in such cases. It may even be sensible to take into account the possibility of period-specific (e.g. month-specific) standard deviations, which can be done by changing the default settings of the **calendarsigma** parameter (see Specifications-X13 section). The value of the **calendarsigma** parameter will have an impact on the method of calculation of the moving standard deviation in the procedure for extreme values detection presented below.

Step 1: Estimating the seasonal component

The seasonal component is estimated by smoothing the SI component separately for each period using a 3×3 moving average, i.e.:

$$\frac{1}{9} \times \left\{ \begin{matrix} 1,0,0,0,0,0,0,0,0,0,0,0, \\ 2,0,0,0,0,0,0,0,0,0,0,0, \\ 3,0,0,0,0,0,0,0,0,0,0,0, \\ 2,0,0,0,0,0,0,0,0,0,0,0, \\ 1,0,0,0,0,0,0,0,0,0,0,0, \end{matrix} \right\}$$

14

Step 2: Normalizing the seasonal factors

The preliminary seasonal factors are normalized in such a way that for one year their average is equal to zero (additive model) or to unity (multiplicative model).

Step 3: Estimating the irregular component

The initial normalized seasonal factors are removed from the Seasonal-Irregular component to provide an estimate of the irregular component.

Step 4: Calculating a moving standard deviation

By default, a moving standard deviation of the irregular component is calculated at five-year intervals. Each standard deviation is associated with the central year used to calculate it. The values in the central year, which in the absolute terms deviate from average by more than the **Usigma** parameter are marked as extreme values and assigned a zero weight. After excluding the extreme values the moving standard deviation is calculated once again.

Step 5: Detecting extreme values and weighting the irregular

The default settings for assigning a weight to each value of irregular component are:

- Values which are more than **Usigma** (2.5, by default) standard deviations away (in the absolute terms) from the 0 (additive) or 1 (multiplicative) are assigned a zero weight;
- Values which are less than 1.5 standard deviations away (in the absolute terms) from the 0 (additive) or 1 (multiplicative) are assigned a full weight (equal to one);
- Values which lie between 1.5 and 2.5 standard deviations away (in the absolute terms) from the 0 (additive) or 1 (multiplicative) are assigned a weight that varies linearly between 0 and 1 depending on their position.

The default boundaries for the detection of the extreme values can be changed with **LSigma** and **USigma** parameters

Step 6: Adjusting extreme values of the seasonal-irregular component

Values of the SI component are considered extreme when a weight less than 1 is assigned to their irregular. Those values are replaced by a weighted average of five values:

- The value itself with its weight;
- The two preceding values, for the same period, having a full weight(if available);
- The next two values, for the same period, having full a weight (if available).

When the four full-weight values are not available, then a simple average of all the values available for the given period is taken.

This general algorithm is used with some modification in parts B and C for detection and replacement of extreme values.

16.1.0.2.4 X-11 tables

The list of tables produced by JDemetra+ is presented below. It is not identical to the output produced by the original X-11 program.

Part A – Preliminary Estimation of Outliers and Calendar Effects.

This part includes prior modifications to the original data made in the RegARIMA part:

- Table A1 – Original series;
- Table A1a – Forecast of Original Series;
- Table A2 – Leap year effect;
- Table A6 – Trading Day effect (1 or 6 variables);
- Table A7 – The Easter effect;
- Table A8 – Total Outlier Effect;
- Table A8i – Additive outlier effect;
- Table A8t – Level shift effect;
- Table A8s – Transitory effect;
- Table A9 – Effect of user-defined regression variables assigned to the seasonally adjusted series or for which the component has not been defined;
- Table 9sa – Effect of user-defined regression variables assigned to the seasonally adjusted series;

- Table9u – Effect of user-defined regression variables for which the component has not been defined.

Part B – Preliminary Estimation of the Time Series Components:

- Table B1 – Original series after adjustment by the RegARIMA model;
- Table B2 – Unmodified Trend (preliminary estimation using composite moving average);
- Table B3 – Unmodified Seasonal – Irregular Component (preliminary estimation);
- Table B4 – Replacement Values for Extreme SI Values;
- Table B5 – Seasonal Component;
- Table B6 – Seasonally Adjusted Series;
- Table B7 – Trend (estimation using Henderson moving average);
- Table B8 – Unmodified Seasonal – Irregular Component;
- Table B9 – Replacement Values for Extreme SI Values;
- Table B10 – Seasonal Component;
- Table B11 – Seasonally Adjusted Series;
- Table B13 – Irregular Component;
- Table B17 – Preliminary Weights for the Irregular;
- Table B20 – Adjustment Values for Extreme Irregulars.

Part C – Final Estimation of Extreme Values and Calendar Effects:

- Table C1 – Modified Raw Series;
- Table C2 – Trend (preliminary estimation using composite moving average);
- Table C4 – Modified Seasonal – Irregular Component;
- Table C5 – Seasonal Component;
- Table C6 – Seasonally Adjusted Series;
- Table C7 – Trend (estimation using Henderson moving average);
- Table C9 – Seasonal – Irregular Component;
- Table C10 – Seasonal Component;
- Table C11 – Seasonally Adjusted Series;
- Table C13 – Irregular Component;

- Table C20 – Adjustment Values for Extreme Irregulars.

Part D – Final Estimation of the Different Components:

- Table D1 – Modified Raw Series;
- Table D2 – Trend (preliminary estimation using composite moving average);
- Table D4 – Modified Seasonal – Irregular Component;
- Table D5 – Seasonal Component;
- Table D6 – Seasonally Adjusted Series;
- Table D7 – Trend (estimation using Henderson moving average);
- Table D8 – Unmodified Seasonal – Irregular Component;
- Table D9 – Replacement Values for Extreme SI Values;
- Table D10 – Final Seasonal Factors;
- Table D10A – Forecast of Final Seasonal Factors;
- Table D11 – Final Seasonally Adjusted Series;
- Table D11A – Forecast of Final Seasonally Adjusted Series;
- Table D12 – Final Trend (estimation using Henderson moving average);
- Table D12A – Forecast of Final Trend Component;
- Table D13 – Final Irregular Component;
- Table D16 – Seasonal and Calendar Effects;
- Table D16A – Forecast of Seasonal and Calendar Component;
- Table D18 – Combined Calendar Effects Factors.

Part E – Components Modified for Large Extreme Values:

- Table E1 – Raw Series Modified for Large Extreme Values;
- Table E2 – SA Series Modified for Large Extreme Values;
- Table E3 – Final Irregular Component Adjusted for Large Extreme Values;
- Table E11 – Robust Estimation of the Final SA Series.

Part F – Quality indicators:

- Table F2A – Changes, in the absolute values, of the principal components;
- Table F2B – Relative contribution of components to changes in the raw series;

- Table F2C – Averages and standard deviations of changes as a function of the time lag;
- Table F2D – Average duration of run;
- Table F2E – I/C ratio for periods span;
- Table F2F – Relative contribution of components to the variance of the stationary part of the original series;
- Table F2G – Autocorrelogram of the irregular component.

16.1.0.3 Filter length choice

A seasonal filter is a weighted average of a moving span of fixed length within a time series that can be used to remove a fixed seasonal pattern. X-13ARIMA-SEATS uses several of these filters, according to the needs of the different stages of the program. As only X-13ARIMA-SEATS allows the user to manually select seasonal filters, this case study can be applied only to the X-13ARIMA-SEATS specifications.

The automatic seasonal adjustment procedure uses the default options to select the most appropriate moving average. However there are occasions when the user will need to specify a different seasonal moving
average to that identified by the program. For example, if the SI values do not closely follow the seasonal component, it may be appropriate to use a shorter moving average. Also the presence of sudden breaks in the seasonal pattern – e.g. due to changes in the methodology – can negatively impact on the automatic selection of the most appropriate seasonal filter. In such cases the usage of short seasonal filters in the selected months or quarters can be considered. Usually, a shorter seasonal filter (3×1) allows seasonality to change very rapidly over time. However, a very short seasonal filter should not normally be used, as it might often lead to large revisions as new data becomes available. If a short filter is to be used it will usually be limited to one month/quarter with a known reason for wanting to capture a rapidly changing seasonality.

In the standard situation one seasonal filter is applied to all individual months/quarters. The estimation of seasonal movements is therefore based on the sample windows of equal lengths for each individual month/quarter (i.e. for each month/quarter the seasonal filter length or the number of years representing the major part of the seasonal filter weights is identical). This approach relies on the assumption that the number of past periods in which the conditions causing seasonal behaviour are sufficiently homogenous is the same in all months/quarters. However, this assumption does not always hold. Seasonal causes may change in one month, while staying the same in others[1]. For instance, seasonal heteroskedasticity might require different filter lengths in different months or quarters.

[1]When the series are non-stationary differentiation is performed before the seasonality tests.

Another interesting example is industrial production in Germany. It can be influenced by school holidays, since many employees have school-age children, which interrupt their working pattern during these school holidays. Consequently, businesses may temporarily suspend or lower production during these periods. Since school holidays do not occur at the same time throughout Germany and their timing varies from year to year in the individual federal states, the effect is not completely captured by seasonal adjustment. And since school holidays are treated as usual working days, these effects are not captured by calendar adjustment either. The majority of school holidays in Germany can take place either in July or in August. This yields higher variances in the irregular component for these months compared to the rest of the year. Therefore, in this case a longer seasonal filter is used for these months to account for this.

Another example might be given by German retail trade. Due to changes in the consumers' behaviour around Christmas – possibly more gifts of money – the seasonal peak in December has become steadily less pronounced. To account for this moving seasonality, shorter seasonal filters in December than during the rest of the year need to be applied.

JDemetra+ offers the options to assign a different seasonal filter length to each period (month or quarter). The program offers these options in the *single spec* mode as well as in the *multispec* mode, albeit they are available only in the *Specifications* window, after a document is created.

16.1.1 M-stats

The details about the measures are given below.

- $M1$ measures the contribution of the irregular component to the total variance. When it is above 1 some changes in outlier correction should be considered.
- $M2$, which is a very similar to $M1$, is calculated on the basis of the contribution of the irregular component to the stationary portion of the variance. When it is above 1, some changes in an outlier correction should be considered.
- $M3$ compares the irregular to the trend taken from a preliminary estimate of the seasonally adjusted series. If this ratio is too large, it is difficult to separate the two components from each other. When it is above 1 some changes in outlier correction should be considered.
- $M4$ tests the randomness of the irregular component. A value above 1 denotes a correlation in the irregular component. In such case a shorter seasonal moving average filter should be considered.
- $M5$ is used to compare the significance of changes in the trend with that in the irregular. When it is above 1 some changes in outlier correction should be considered.

- $M6$ checks the SI (seasonal – irregular components ratio). If annual changes in the irregular component are too small in relation to the annual changes in the seasonal component, the 3×5 seasonal filter used for the estimation of the seasonal component is not flexible enough to follow the seasonal movement. In such case a longer seasonal moving average filter should be considered. It should be stressed that $M6$ is calculated only if the 3×5 filter has been applied in the model.
- $M7$ is the combined test for the presence of an identifiable seasonality. The test compares the relative contribution of stable and moving seasonality[2].
- $M8$ to $M11$ measure if the movements due to the short-term quasi-random variations and movements due to the long-term changes are not changing too much over the years. If the changes are too strong then the seasonal factors could be erroneous. In such case a correction for a seasonal break or the change of the seasonal filter should be considered.

The Q statistic is a composite indicator calculated from the M statistics.

Edit : problem with tables display

$$Q = \frac{10M1 + 11M2 + 10M3 + 8M4 + 11M5 + 10M6 + 18M7 + 7M8 + 7M9 + 4M10 + 4M11}{100}$$

[1]

$Q = Q - M2$ (also called $Q2$) is the Q statistic for which the $M2$ statistics was excluded from the formula, i.e.:

$$Q - M2 = \frac{10M1 + 10M3 + 8M4 + 11M5 + 10M6 + 18M7 + 7M8 + 7M9 + 4M10 + 4M11}{89}$$

[2]

If a time series does not cover at least 6 years, the $M8$, $M9$, $M10$ and $M11$ statistics cannot be calculated. In this case the Q statistic is computed as:

$$Q = \frac{14M1 + 15M2 + 10M3 + 8M4 + 11M5 + 10M6 + 32M7}{100}$$

The model has a satisfactory quality if the Q statistic is lower than 1.

The tables displayed in the *Quality measures* $\rightarrow$ *Details* node correspond to the F-set of tables produced by the original X-11 algorithm. To facilitate the analysis of the results, the numbers and the names of the tables are given under each table following the convention used in LADIRAY, D., and QUENNEVILLE, B. (1999).

[2] See section Combined seasonality tests.

16.1.2 Detailed tables

The first table presents the average percent change without regard to sign of the percent changes (multiplicative model) or average differences (additive model) over several periods (from 1 to 12 for a monthly series, from 1 to 4 for a quarterly series) for the following series:

- O – Original series (Table A1);
- CI – Final seasonally adjusted series (Table D11);
- I – Final irregular component (Table D13);
- C – Final trend (Table D12);
- S – Final seasonal factors (Table D10);
- P – Preliminary adjustment coefficients, i.e. regressors estimated by the RegARIMA model (Table A2);
- $TD\&H$ – Final calendar component (Tables A6 and A7);
- Mod.O – Original series adjusted for extreme values (Table E1);
- Mod.CI – Final seasonally adjusted series corrected for extreme values (Table E2);
- Mod.I – Final irregular component adjusted for extreme values (Table E3).

In the case of an additive decomposition, for each component the average absolute changes over several periods are calculated as[3]:

$$\text{Component}_d = \frac{1}{n-d}\sum_{t=d+1}^{n} |Table_t - Table_{t-d}|$$

[4]

where:

d – time lag in periods (from a monthly time series d varies from to 4 or from 1 to 12);

n – total number of observations per period;

Component – the name of the component;

Table – the name of the table that corresponds to the component.

Table F2A – changes, in the absolute values, of the principal components

[3]For the multiplicative decomposition the following formula is used:

$$\text{Component}_d = \frac{1}{n-d}\sum_{t=d+1}^{n} |\frac{\text{Table}_t}{\text{Table}_{t-d}} - 1|$$

.

Average percent change without regard to sign over the indicated span

Span	O	CI	I	C	S	P	TD&H	Mod.O	Mod.CI	Mod.I
1	7,50	3,81	3,49	1,42	6,99	0,00	0,00	7,75	3,57	3,29
2	5,33	4,88	3,90	2,88	3,57	0,00	0,00	5,40	4,61	3,55
3	8,23	5,75	3,74	4,39	7,16	0,00	0,00	8,53	5,50	3,39
4	6,36	6,75	3,76	5,94	0,00	0,00	0,00	6,74	6,74	3,56

Figure 16.2: Text

Next, Table F2B of relative contributions of the different components to the differences (additive model) or percent changes (multiplicative model) in the original series is displayed. They express the relative importance of the changes in each component. Assuming that the components are independent, the following relation is valid:

$$O_d^2 \approx C_d^2 + S_d^2 + I_d^2 + P_d^2 + TD\&H_d^2$$

. [5]

In order to simplify the analysis, the approximation can be replaced by the following equation:

$$O_d^{*2} = C_d^2 + S_d^2 + I_d^2 + P_d^2 + TD\&H_d^2$$

. [6]

The notation is the same as for Table F2A. The column Total denotes total changes in the raw time series.

Data presented in Table F2B indicate the relative contribution of each component to the percent changes (differences) in the original series over each span, and are calculated as:

$\frac{I_d^2}{O_d^{*2}}$,

$\frac{C_d^2}{O_d^{*2}}$,

$\frac{S_d^2}{O_d^{*2}}$,

$\frac{P_d^2}{O_d^{*2}}$

and $\frac{TD\&H_d^2}{O_d^{*2}}$

where: $O_d^{*2} = I_d^2 + C_d^2 + S_d^2 + P_d^2 + TD\&H_d^2$.

The last column presents the *Ratio* calculated as:

$100 \times \frac{O_d^{*2}}{O_d^2}$,

which is an indicator of how well the approximation

$$(O_d^*)^2 \approx O_d^2$$

holds.

Relative contributions to the variance of the percent change in the components of the original series

Span	I	C	S	P	TD&H	Total	Ratio
1	17,53	3,27	79,20	0,00	0,00	100,00	102,79
2	37,38	24,71	37,91	0,00	0,00	100,00	115,35
3	13,97	23,47	62,56	0,00	0,00	100,00	112,79
4	26,47	73,53	0,00	0,00	0,00	100,00	105,49

Figure 16.3: Text

Table F2B – relative contribution of components to changes in the raw series

When an additive decomposition is used, Table F2C presents the average and standard deviation of changes calculated for each time lag d, taking into consideration the sign of the changes of the raw series and its components. In case of a multiplicative decomposition the respective table shows the average percent differences and related standard deviations.

Average percent change with regard to sign and standard deviation over indicated span

	O		I		C		S		CI	
Span	Avg	S.D.	Avg	S.D.	Avg	S.D.	Avg	S.D.	Avg	S.D.
1	1,97	8,67	0,05	3,73	1,41	0,48	0,53	8,01	1,46	3,81
2	3,19	5,72	0,15	4,48	2,86	0,97	0,24	4,42	3,02	4,72
3	4,97	9,47	0,09	4,52	4,36	1,44	0,55	8,30	4,46	4,97
4	5,93	3,81	0,10	4,32	5,90	1,90	0,00	0,00	6,01	5,06

Figure 16.4: Text

Table F2C – Averages and standard deviations of changes as a function of the time lag

Average duration of run is an average number of consecutive monthly (or quarterly) changes in the same direction (no change is counted as a change in the same direction as the preceding change). JDemetra+ displays this indicator for the seasonally adjusted series, for the trend and for the irregular component.

Average duration of run

CI	8.44
I	1.31
C	15.20

Figure 16.5: Text

Table F2D – Average duration of run

The $\frac{I}{C}$ ratios for each value of time lag d, presented in Table F2E, are computed on a basis of the data in Table F2A.

I/C Ratio for indicated span

1	**0.150**
2	0.052
3	0.039
4	0.031

I/C Ratio: 0.314

Figure 16.6: Text

Table F2E – $\frac{\mathbf{I}}{\mathbf{C}}$ ratio for periods span

The relative contribution of components to the variance of the stationary part of the original series is calculated for the irregular component (I), trend made stationary[4] (C), seasonal component (S) and calendar effects (TD&H). The short description of the calculation method is given in LADIRAY, D., and QUENNEVILLE, B. (1999).

Relative contribution of the components to the stationary portion of the variance in the original series.

I	0.01
C	99.56
S	0.15
P	0.00
TD&H	0.00
Total	99.72

Figure 16.7: Text

Table F2F – Relative contribution of components to the variance of the stationary part of the original series

The last table shows the autocorrelogram of the irregular component from Table D13. In the case of multiplicative decomposition it is calculated for time lags between 1 and the number of periods per year +2 using the formula[5]:

[4]The component is estimated by extracting a linear trend from the trend component presented in Table D12.
[5]For the additive decomposition the formula is:

$$Corr_k I_t = \frac{\sum_{t=k+1}^{N} (I_t \times I_{t-k})}{\sum_{t=1}^{N} (I_t)^2}$$

$$\mathrm{Corr}_k I = \frac{\sum_{t=k+1}^{N} (I_t - 1)(I_{t-k} - 1)}{\sum_{t=1}^{N} (I_t - 1)^2}]$$

[7]

where N is number of observations in the time series and k the lag.

Autocorrelation of the irregular

1	-0.601
2	0.200
3	0.019
4	-0.147
5	0.187
6	-0.138

Figure 16.8: Text

Table F2G – Autocorrelation of the irregular component

The Cochran test is design to identify the heterogeneity of a series of variances. X-13-ARIMA-SEATS uses this test in the extreme value detection procedure to check if the irregular component is heteroskedastic. In this procedure the standard errors of the irregular component are used for an identification of extreme values. If the null hypothesis that for all the periods (months, quarters) the variances of the irregular component are identical is rejected, the standard errors will be computed separately for each period (in case the option *Calendarsigma*=**signif** has been selected).

Heteroskedasticity (Cochran test on equal variances within each period)

Test statistic	Critical value (5% level)	Decision
0.1303	0.15	Null hypothesis is not rejected.

Figure 16.9: Text

Cochran test

For each i^{th} month we will be looking at the mean annual changes for each component by calculating:

$$\overline{S}_i = \frac{1}{N_i - 1} \sum_{t=2}^{N_i} |S_{i,t} - S_{i,t-1}|$$

and

$$\overline{I}_i = \frac{1}{N_i - 1} \sum_{t=2}^{N_i} |I_{i,t} - I_{i,t-1}|$$

,

where N_i refers to the number of months i in the data, and the moving seasonality ratio of month i:

$$MSR_i = \frac{\overline{I}_i}{\overline{\overline{S}}_i}$$

These ratios are published in Table D9A in X13ARIMA-SEATS software. In JDemetra+ they are presented in the details of the quality measures.

The Moving Seasonality Ratio (MSR) is used to measure the amount of noise in the Seasonal-Irregular component. By studying these values, the user can select for each period the seasonal filter that is the most suitable given the noisiness of the series.

Moving Seasonality Ratios (MSR)

Period	I	S	MSR
1	0.0597	0.0211	2.8292
2	0.0808	0.0135	5.9850
3	0.0767	0.0139	5.5038
4	0.0777	0.0262	2.9640

Figure 16.10: Text

Table D9a – Moving seasonality ratios

17 STL: Local regression decomposition

18 SEATS decomposition

18.1 Introduction

SEATS is a program for estimating unobserved components in a time series. It follows the ARIMA-model-based (AMB) method, developed from the work of CLEVELAND, W.P., and TIAO, G.C. (1976), BURMAN, J.P. (1980), HILLMER, S.C., and TIAO, G.C. (1982), BELL, W.R., and HILLMER, S.C. (1984) and MARAVALL, A., and PIERCE, D.A. (1987).

In JDemetra+ the input for the model based signal extraction procedure is always provided by TRAMO and includes the original series y_t, the linearized series x_t (i.e. the original series y_t with the deterministic effects removed), the ARIMA model for the stochastic (linearized) time series x_t and the deterministic effects (calendar effects, outliers and other regression variable effects)[1]. SEATS decomposes the linearized series (and the ARIMA model) into trend, seasonal, transitory and irregular components, provides forecasts for these components, together with the associated standard errors, and finally assign the deterministic effects to each component yielding the *final* components[2]. The Minimum Mean Square Error (MMSE) estimators of the components are computed with a Wiener-Kolmogorov filter applied to the finite series extended with forecasts and backcasts[3].

18.2 ARIMA modellling of the input series

One of the fundamental assumptions made by SEATS is that the linearized time series x_t follows the ARIMA model

$$\phi(B)\delta\left(B\right)x_t = \theta(B)a_t$$

[1]

where:

B – the backshift operator ($Bx_t = x_{t-1}$);

[1] In the original software SEATS can be used either with TRAMO, operating on the input received from the latter, or alone, fitting an ARIMA model to the series.

[2] GÓMEZ, V., and MARAVALL, A. (1998).

[3] GÓMEZ, V., and MARAVALL, A. (1997).

$\delta(B)$ – a non-stationary autoregressive (AR) polynomial in B (unit roots);

$\theta(B)$ – an invertible moving average (MA) polynomial in B and in B^S, which can be expressed in the multiplicative form

$(1+\vartheta_1 B+\ldots+\vartheta q B^q)\left(\ 1+\Theta 1 B^s+\ldots+\Theta_Q B^{\mathrm{sQ}}\right)$;

$\phi(B)$ – a stationary autoregressive (AR) polynomial in B and in B^S containing regular and seasonal unit roots, with s representing the number of observations per year;

a_t – a white-noise variable with the variance $V(a)$.

It should be noted that the stochastic time series can be predicted using its past observations and making an error. The variable a_t, which is assumed to be white noise, is the fundamental *innovation* to the series at time t, that is the part that cannot be predicted based on the past history of the series.

Denoting $\varphi(B)=\phi(B)\,\delta(B)$, [1] can be written in a more concise form as

$$\varphi(B)\,x_t=\theta(B)a_t$$

, [2]

where $\varphi(B)$ contains both the stationary and the nonstationary roots.

18.3 Derivation of the models for the components

Let us consider the additive decomposition model

$$x_t=\sum_{i=1}^{k} x_{\mathrm{it}}$$

, [3]

where i refers to the orthogonal components: trend, seasonal, transitory or irregular. Apart from the irregular component, supposed to be a white noise, it is assumed that each component follows the ARIMA model which can be represented, using the notation of [2] , as:

$$\varphi_i(B)\ \ x_{\mathrm{it}}=\theta_i(B)a_{\mathrm{it}}$$

, [4]

where $\varphi_i(B)=\phi_i(B)\,\delta_i(B)$, x_{it} is the *i-th* unobserved component, $\varphi_i(B)$ and $\theta_i(B)$ are finite polynomials of order p_i and q_i, respectively, and a_{it}, the disturbance associated with such component, is a white noise process with zero mean and constant variance $V(a_i)$ and a_{it}

and a_{jt} are not correlated for $i \neq j$ and for any t.. These disturbances are functions of the innovations in the series and are called ”pseudo-innovations” in the literature concerning the AMB decomposition as they refer to the components that are never observed [4]. In the JDemetra+ documentation the term ”innovations” is used to refer to the ”pseudo-innovations”.

The following assumptions hold for [4] . For each i the polynomials $\phi_i\left(B\right)$, $\delta_i\left(B\right)$ and $\theta_i(B)$ are prime and of finite order. The roots of $\delta_i\left(B\right)$ lies on the unit circle; those of $\phi_i\left(B\right)$ lie outside, while all the roots of $\theta_i\left(B\right)$ are on or outside the unit circle. This means that nonstationary and noninvertible components are allowed. Since different roots of the AR polynomial induce peaks in the spectrum[5] of the series at different frequencies, and given that different components are associated with the spectral peaks for different frequencies, it is assumed that for $i \neq j$ the polynomials $\phi_i\left(B\right)$ and $\phi_j\left(B\right)$ do not share any common root (they are coprime). Finally, it is assumed that the polynomials $\theta_i\left(B\right),\ i = 1,\ldots,k$ are prime share no unit root in common, guaranteeing the invertibility of the overall series. In fact, since the unit root of $\theta_i\left(B\right)$ induce a spectral zero, when the polynomials $\theta_i\left(B\right),\ i = 1,\ldots,k$ share no unit root in common, there is no frequency for which all component spectra become zero[6].

Since aggregation of ARIMA models yields ARIMA models, the series x_t will also follow an ARIMA model, as in [2] , and consequently the following identity can be derived:

$$\frac{\theta(B)}{\varphi(B)}a_t = \sum_{i=1}^{k}\frac{\theta_i(B)}{\varphi_i(B)}a_{\mathrm{it}}$$

. [5]

In the ARIMA model based approach implemented in SEATS, the ARIMA model identified and estimated for the observed series x_t is decomposed to derive the models for the components. In particular, the AR polynomials for the components, $\varphi_i\left(B\right)$, are easily derived through the factorization of the AR polynomial $\varphi\left(B\right)$:

$$\varphi\left(B\right) = \prod_{i=1}^{k}\varphi_i\left(B\right)$$

, [6]

while the MA polynomials for the components, together with the innovation variances $V(a_i)$, cannot simply be obtained through the relationship:

$$\theta(B)a_t = \sum_{i=1}^{k}\varphi_{\mathrm{ni}}\left(B\right)\theta_i(B)a_{\mathrm{it}}$$

[4] GÓMEZ, V., and MARAVALL, A. (2001a).

[5] For description of the spectrum see section Spectral Analysis.

[6] MARAVALL, A. (1995).

, [7]

where $\varphi_{\text{ni}}(B)$ is the product of all $\varphi_j(B)$, $j = 1, \ldots, k$, except from $\varphi_i(B)$. Further assumptions are therefore needed to cope with the underidentification problem: i) $p_i \geq q_i$ and ii) the canonical decomposition, i.e. the decomposition that allocate all additive white noise to the irregular component (yielding noninvertible components except the irregular).

To understand how SEATS factorizes the AR polynomials, first a concept of a root will be explored[7].

The equation [2] can be expressed as:

$$\psi^{-1}(B)x_t = a_t(1 + \varphi_1 B + \ldots \varphi_p B^p)x_t = (1 + \theta_1 B + \ldots \theta_q B^q)a_t$$

, [8]

Let us now consider [2] in the inverted form:

$$\theta(B)\, y_t = \varphi(B)a_t$$

, [9]

If both sides of [8] are multiplied by x_{t-k} with $k > q$, and expectations are taken, the right hand side of the equation vanishes and the left hand side becomes:

$$\varphi(B)\gamma_k = \gamma_k + \varphi_1\gamma_{k-1} + \ldots \varphi_p\gamma_{k-p} = 0$$

, [10]

where B operates on the subindex k.

The autocorrelation function γ_k is a solution of [10] with the characteristic equation:

$$z^p + \varphi_1 z^{p-1} + \ldots \varphi_{p-1} z + \varphi_p = 0$$

. [11]

If z_1,..., z_p are the roots of [11] , the solutions of [10] can be expressed as:

$\gamma_k = \sum_{i=1}^{p} z_i^k$, [12]

and will converge to zero as $k \to \infty$ when $|r_i| < 1$, $i = 1, \ldots, p$. From [10] and [12] it can be noticed that $z_1 = B_i^{-1}$, meaning that z_1,..., z_p are the inverses of the roots $B_1, \ldots, B_p$ of the polynomial $\varphi(B)$. The convergence of γ_k implies that the roots of the $\varphi(B)$ are larger than 1 in modulus (lie outside the unit circle). Therefore, from the equation

[7] Description based on KAISER, R., and MARAVALL, A. (2000) and MARAVALL, A. (2008c).

$$\varphi(B)^{-1} = \frac{1}{(1 - z_1) \ldots (1 - z_1)}$$

[13]

it can be derived that $\varphi(B)^{-1}$ is convergent and all its inverse roots are less than 1 in modulus.

Equation [11] has real and complex roots (solutions). Complex number $x = a + bi$, with a and b both real numbers, can be represented as $x = r\,(cos(\omega) + i\ sin(\omega))$, where i is the imaginary unit ($i^2 = -1$), r is the modulus of x, that is $r = |x| = \sqrt{a^2 + b^2}$ and ω is the argument (frequency). When roots are complex, they are always in pairs of complex conjugates. The representation of the complex number $x = a + bi$ has a geometric interpretation in the complex plane established by the real axis and the orthogonal imaginary axis.

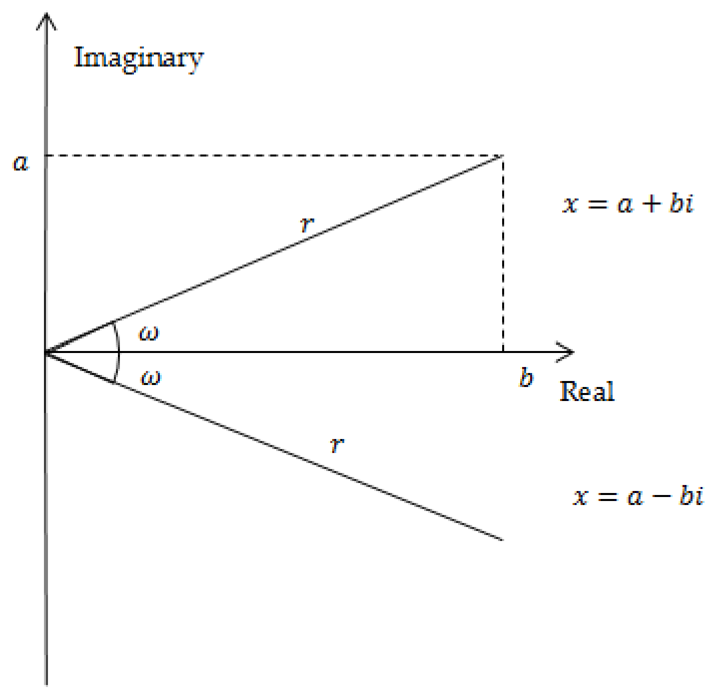

Figure 18.1: Text

Geometric representation of a complex number and of its conjugate

Representing the roots of the characteristic equation [11] in the complex plane enhances understanding how they are allocated to the components. When the modulus r of the roots in z are greater than 1 (i.e. modulus of the roots in $\varphi(B) < 1$), the solution of the characteristic equation has a systematic explosive process, which means that the impact of the given impulse on the time series is more and more pronounced in time. This behaviour is not in line with the developments that can be identified in actual economic series. Therefore, the models estimated by TRAMO-SEATS (and X-13ARIMA-SEATS) have never inverse roots in B with modulus greater than 1.

The characteristic equations associated with the regular and the seasonal differences have roots in $\varphi(B)$ with modulus $r = 1$. They are called non-stationary roots and can be represented on the unit circle. Let us consider the seasonal differencing operator applied to a quarterly time series $(1 - B^4)$. Its characteristic equation is $(z^4 - 1) = 0$ with solutions given by $z = \sqrt[4]{1}$, i.e. $z_{1,2} = \pm 1$ and $z_{3,4} = \pm i1$. The first two solutions are real and the last two are complex conjugates. They are represented by the black points on the unit circle on the figure below.

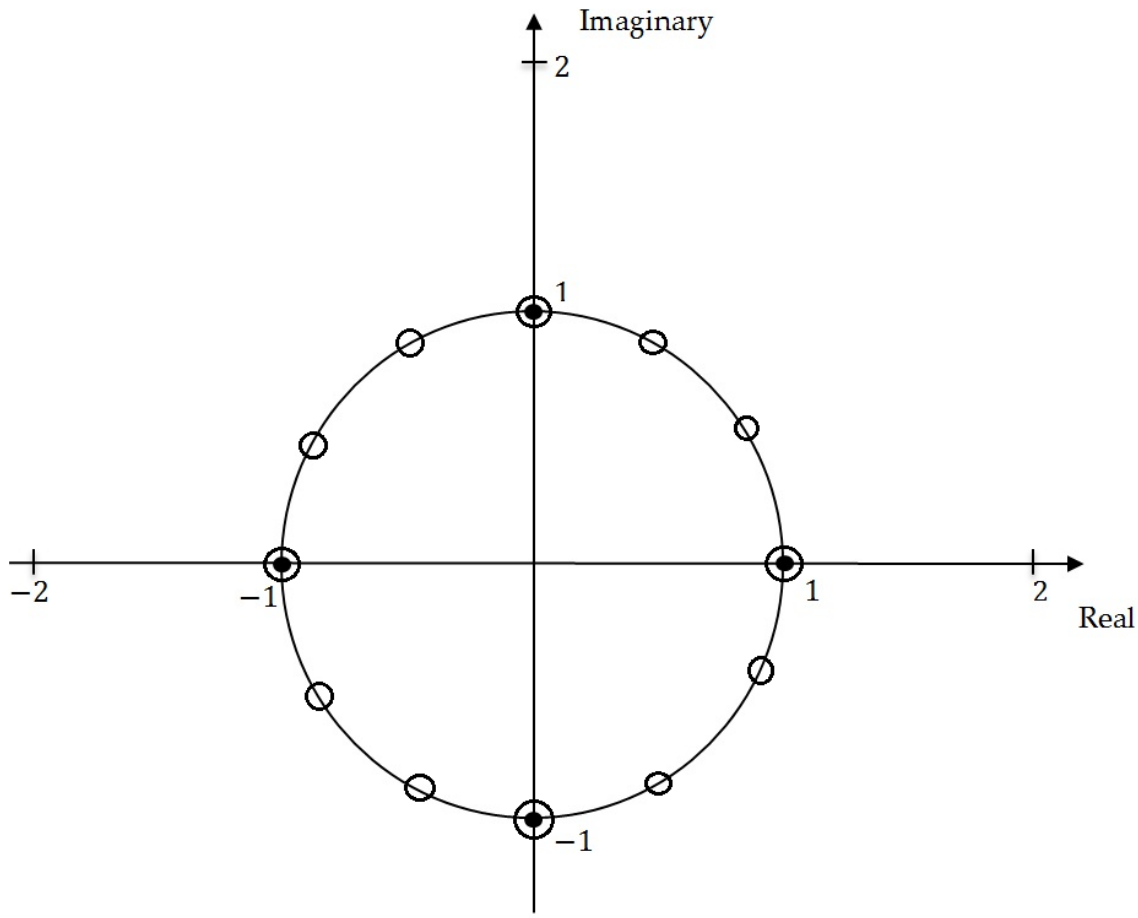

Figure 18.2: Text

Unit roots on the unit circle

For the seasonal differencing operator $(1 - B^{12})$ applied to the monthly time series the characteristic equation $(z^{12} - 1) = 0$ has twelve non-stationary solutions given by $z = \sqrt[12]{1}$: two real and ten complex conjugates, represented by the white circles in unit roots figure above.

The complex conjugates roots generate the periodic movements of the type:

$$z_t = A^t \cos\left(\omega t + W\right).$$

[14]

where:

A – amplitude;

ω – angular frequency (in radians);

W – phase (angle at $t = 0$).

The frequency f, i.e. the number of cycles per unit time, is $\frac{\omega}{2\pi}$. If it is multiplied by s, the number of observations per year, the number of cycles completed in one year is derived. The period of function [14] , denoted by τ, is the number of units of time (months/quarters) it takes for a full circle to be completed.

For quarterly series the seasonal movements are produced by complex conjugates roots with angular frequencies at $\frac{\pi}{2}$ (one cycle per year) and π (two cycles per year). The corresponding number of cycles per year and the length of the movements are presented in the table below.

Seasonal frequencies for a quarterly time series

{: .table .table-style}

Angular frequency (ω)	**Frequency (cycles per unit time) (f)**	**Cycles per year**	**Length of the movement measured in quarters (τ)**
$\frac{\pi}{2}$	0.25	1	4
π	0.5	2	2

For monthly time series the seasonal movements are produced by complex conjugates roots at the angular frequencies: $\frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}$ and π. The corresponding number of cycles per year and the length of the movements are presented in the table below: Seasonal frequencies for a monthly time series.

Seasonal frequencies for a monthly time series

{: .table .table-style}

Angular frequency (ω)	**Frequency (cycles per unit time) (f)**	**Cycles per year**	**Length of the movement measured in months (τ)**
$\frac{\pi}{6}$	0.083	1	12
$\frac{\pi}{3}$	0.167	2	6
$\frac{\pi}{2}$	0.250	3	4
$\frac{2\pi}{3}$	0.333	4	3
$\frac{5\pi}{6}$	0.417	5	2.4
π	0.500	6	2

In JDemetra+ SEATS assigns the roots of the AR full polynomial to the components according to their associated modulus and frequency, i.e.:[8]

[8] For details see MARAVALL, A., CAPORELLO, G., PÉREZ, D., and LÓPEZ, R. (2014).

- Roots of $(1-B)^d$ are assigned to trend component.
- Roots of $(1-B^s)^{d_s} = ((1-B)(1+B+...+B^{s-1}))^{d_s}$ are assigned to the trend component (root of $(1-B)^{d_s}$) and to the seasonal component (roots of $(1+B+...+B^{s-1})^{d_s}$).
- When the modulus of the inverse of a real positive root of $\varphi(B)$ is greater than k or equal to k, where k is the threshold value controlled by the *Trend boundary* parameter(in the original SEATS it is controlled by *rmod*)[9], then the root is assigned to the trend component. Otherwise it is assigned to the transitory component.
- Real negative inverse roots of $\ _p(B)$ associated with the seasonal two-period cycle are assigned to the seasonal component if their modulus is greater than k, where k is the threshold value controlled by the *Seasonal boundary* and the *Seas. boundary (unique)* parameters. Otherwise they are assigned to the transitory component.
- Complex roots, for which the argument (angular frequency) is close enough to the seasonal frequency are assigned to the seasonal component. Closeness is controlled by the *Seasonal tolerance* and *Seasonal tolerance (unique)* parameters (in the original SEATS it is controlled by *epsphi*). Otherwise they are assigned to the transitory component.
- If d_s (seasonal differencing order) is present and Bphi < 0 (Bphi is the estimate of the seasonal autoregressive parameter), the real positive inverse root is assigned to the trend component and the other $(s-1)$ inverse roots are assigned to the seasonal component. When $d_s = 0$, the root is assigned to the seasonal when Bphi < -0.2 and/or the overall test for seasonality indicates presence of seasonality. Otherwise it goes to the transitory component. Also, when Bphi > 0, roots are assigned to the transitory component.

For further details about JDemetra+ parameters see section TramoSeats.

It should be highlighted that when $Q > P$, where Q and P denote the orders of the polynomials $\varphi(B)$ and $\theta(B)$, the SEATS decomposition yields a pure MA $(Q-P)$ component (hence transitory). In this case the transitory component will appear even when there is no AR factor allocated to it.

Once these rules are applied, the factorization of the AR polynomial presented by [2] yields to the identification of the AR polynomials for the components which contain, respectively, the AR roots associated with the trend component, the seasonal component and the transitory component.[10]

Then with the partial fraction expansion the spectrum of the final components are obtained.

[9] In JDemetra+ this argument is called *Trend boundary*.

[10] The AR roots close to or at the trading day frequency generates a stochastic trading day component. A stochastic trading day component is always modelled as a stationary ARMA(2,2), where the AR part contains the roots close to the TD frequency, and the MA(2) is obtained from the model decomposition (MARAVALL, A., and PÉREZ, D. (2011)). This component, estimated by SEATS, is not implemented by the current version of JDemetra+.

For example, the Airline model for a monthly time series:

$$(1-B)(1-B^{12})x_t = (1+\theta_1 B)(1+\Theta_1 B^{12})\ a_t$$

, [15]

is decomposed by SEATS into the model for the trend component:

$$(1-B)(1-B)c_t = (1+\theta_{c,1}B+\theta_{c,2}B^2)a_{c,t}$$

, [16]

and the model for the seasonal component:

$$(1+B+...+B^{11})\, s_t = \left(1+\theta_{s,1}B+...+\theta_{s,11}B^{11}\right) a_{s,t},$$

[17]

As a result, the Airline model is decomposed as follows:

$$\frac{(1+\theta_1 B)(1+\Theta_1 B^{12})}{(1-B)(1-B)}a_t = \frac{\left(1+\theta_{s,1}B+...+\theta_{s,11}B^{11}\right)}{(1+B+...+B^{11})}a_{s,t} + \frac{(1+\theta_{c,1}B+\theta_{c,2}B^2)}{(1-B)(1-B)}a_{c,t} + u_t$$

. [18]

The transitory component is not present in this case and the irregular component is the white noise.

The partial fractions decomposition is performed in a frequency domain. In essence, it consists in portioning of the pseudo-spectrum[11] of x_t into additive spectra of the components. When the AMB decomposition of the ARIMA model results in the non-negative spectra for all components, the decomposition is called admissible[12]. In such case an infinite number of admissible decompositions exists, i.e. decompositions that yield the non-negative spectra of all components. Therefore, the MA polynomials and the innovation variances cannot be yet identified from the model of x_t. As sketched above, to solve this underidentification problem and identify a unique decomposition, it is assumed that for each component the order of the MA polynomial is no greater than the order of the AR polynomial and the canonical solution

[11] The term pseudo-spectrum is used for a non-stationary time series, while the term spectrum is used for a stationary time series.

[12] If the ARIMA model estimated in TRAMO does not accept an admissible decomposition, SEATS replaces it with a decomposable approximation. The modified model is therefore used to decompose the series. There are also other rare situations when the ARIMA model chosen by TRAMO is changed by SEATS. It happens when, for example, the ARIMA models generate unstable seasonality or produce a senseless decomposition. Such examples are discussed by MARAVALL, A. (2009).

of S.C. Hillmer and G.C. Tiao is applied[13], i.e. all additive white noise is added to the irregular component As a consequence all components derived from the canonical decomposition, except from the irregular, have a spectral minimum of zero and are thus noninvertible[14]. Given the stochastic features of the series, it can be shown by that the canonical decomposition produces as stable as possible trend and seasonal components since it maximizes the variance of the irregular and minimizes the variance of the other components[15]. However, there is a price to be paid as canonical components can produce larger revisions in the preliminary estimators of the component[16] than any other admissible decomposition.

The figure below represents the pseudo-spectrum for the canonical trend and an admissible trend.

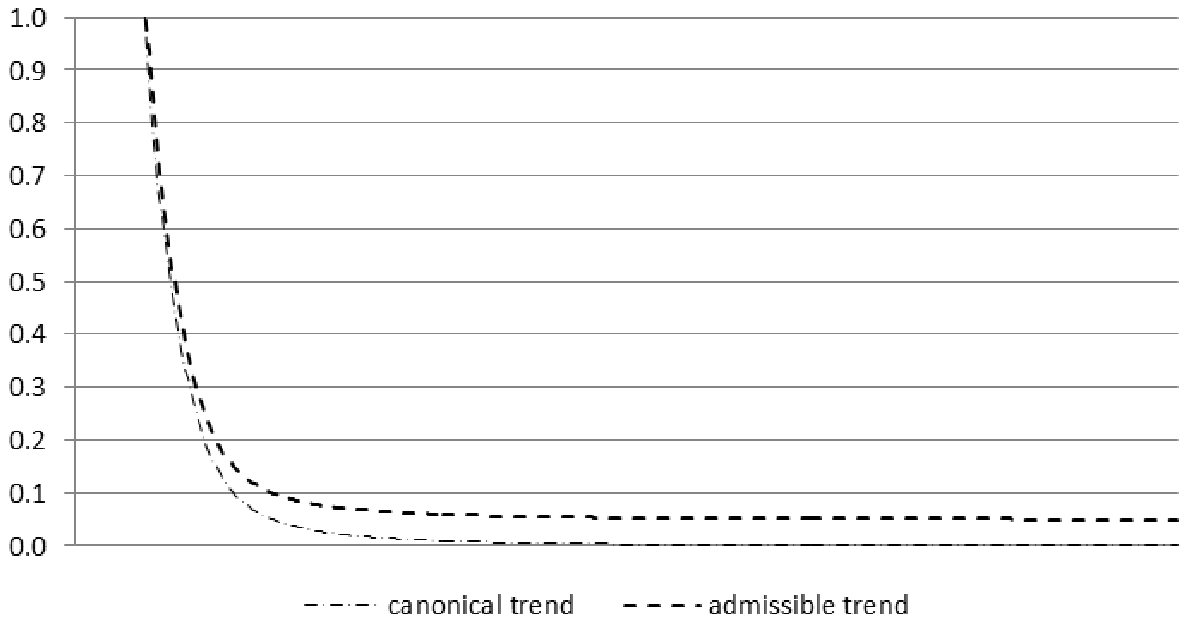

Figure 18.3: Text

A comparison of canonical trend and admissible trend

A pseudo-spectrum is denoted by $g_i(\omega)$, where ω represents the angular frequency. The pseudo-spectrum of x_{it} is defined as the Fourier transform of ACGF of x_t which is expressed as:

$$\frac{\psi_i\left(B\right)\psi_i\left(F\right)}{\delta_i\left(B\right)\delta_i\left(F\right)}V(a_i)$$

[13]HILLMER, S.C., and TIAO, G.C. (1982).
[14]GÓMEZ, V., and MARAVALL, A. (2001a).
[15]HILLMER, S.C., and TIAO, G.C. (1982).
[16]MARAVALL, A. (1986).

, [19]

where:

$\psi_i\left(F\right) = \frac{\theta_i(F)}{\phi_i(F)}$

$\psi_i\left(B\right) = \frac{\theta_i(B)}{\phi_i(B)}$

B is the backward operator,

F is the forward operator.

A pseudo-spectrum for a monthly time series x_t is presented in the figure below: The pseudo-spectrum for a monthly series. The frequency $\omega = 0$ is associated with the trend, frequencies in the range $[0+\epsilon_1,\ \frac{\pi}{6}-\epsilon_2]$ with $[0+\epsilon_1,\ \frac{\pi}{6}-\epsilon_2]$ $\epsilon_1,\ \epsilon_2 > 0$ and $\epsilon_1 <\ \frac{\pi}{6}-\epsilon_2$ are usually associated with the business-cycle and correspond to a period longer than a year and bounded[17]. The frequencies in the range $[\frac{\pi}{6},\ \pi]$ are associated with the short term movements, whose cycle is completed in less than a year. If a series contains an important periodic component, its spectrum reveals a peak around the corresponding frequency and in the ARIMA model it is captured by an AR root. In the example below spectral peaks occur at the frequency $\omega = 0$ and at the seasonal frequencies ($\frac{\pi}{6}$, $\frac{2\pi}{6}$, $\frac{3\pi}{6}$, $\frac{4\pi}{6}$, $\frac{5\pi}{6}$, π). [18]

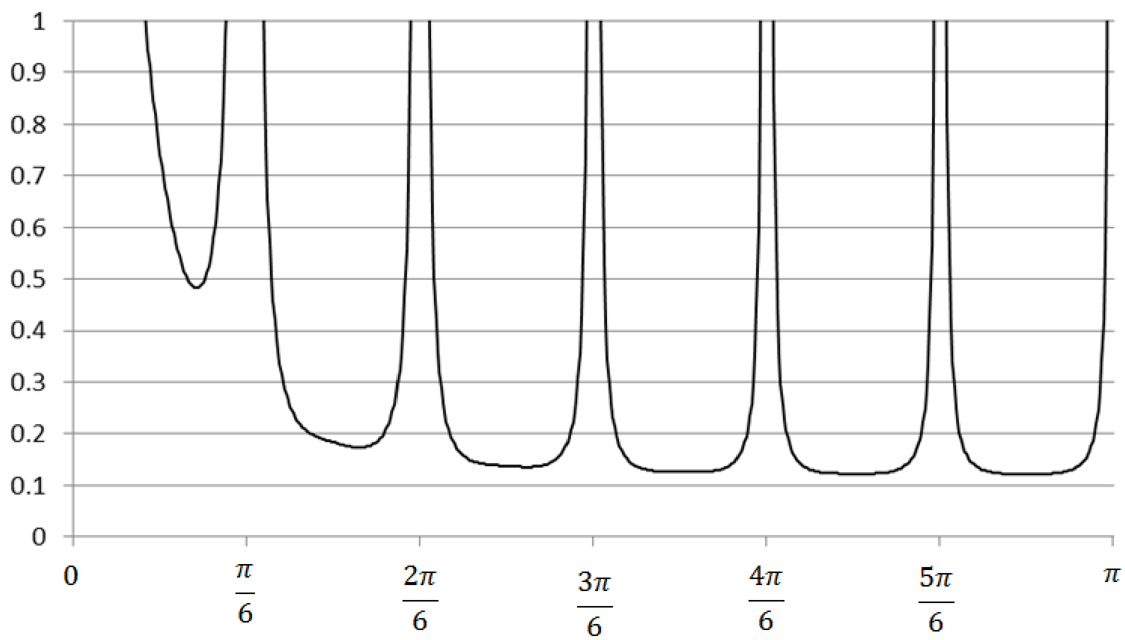

Figure 18.4: Text

The pseudo-spectrum for a monthly series

[17]Ibid.

[18]KAISER, R., and MARAVALL, A. (2000).

In the decomposition procedure, the pseudo-spectrum of the time series x_t is divided into the spectra of its components (in the example figure below, four components were obtained).

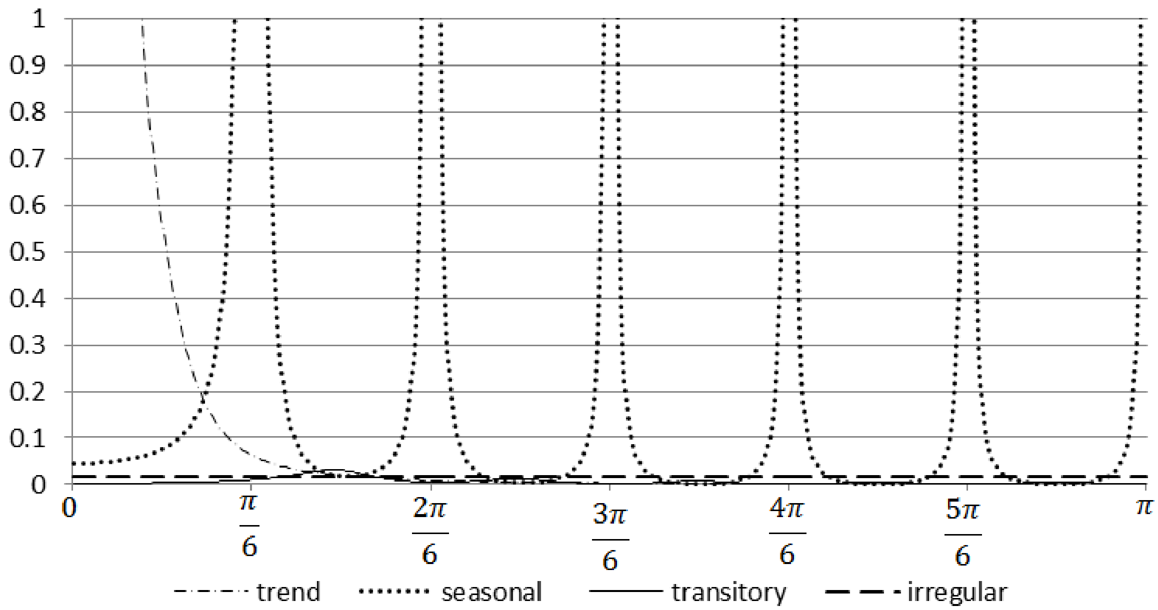

Figure 18.5: Text

The pseudo-spectra for the components

18.4 Estimation of the components with the Wiener-Kolmogorow filter

The various components are estimated using Wiener-Kolmogorow (WK) filters. JDemetra+ includes three options to estimate the WK filter, namely *Burman*, *KalmanSmoother* and *MCElroyMatrix*[19]. Here the first of abovementioned options, proposed by BURMAN, J.P. (1980) will be explained.

The estimation procedure and the properties of the WK filter are easier to explain with a two-component model. Let the seasonally adjusted series (s_t) be the signal of interest and the seasonal component (n_t) be the remainder, "the noise". The series is given by the model [2] and from [4] the models for theoretical components are:

$$\varphi_s(B)s_t = \theta_s(B)a_{\mathrm{st}}$$

[19] The choice of the estimation method is controlled by the *Method* parameter, explained in the SEATS specification section.

[20]

and

$$\varphi_n(B)n_t = \theta_n(B)a_{\mathrm{nt}}$$

. [21]

From [6] and [7] it is clear that $\varphi\left(B\right) = \varphi_s(B)\varphi_n(B)$ and $\theta\left(B\right)a_t = \theta_s(B)a_{\mathrm{st}} + \theta_n(B)a_{\mathrm{nt}}$.

As the time series components are never observed, their estimators have to be used. Let us note X_T an infinite realization of the time series x_t. SEATS computes the Minimum Mean Square Error (MMSE) estimator of s_t, e.g. the estimator

$$\hat{s}_t$$

that minimizes

$$E[\left(s_t - \hat{s}_t\right)^2|X_T\Big)]$$

. Under the normality assumption

$$\hat{s}_{t|T}$$

is also equal to the conditional expectation

$$E\left(s_t|X_T\right)$$

, so it can be presented as a linear function of the elements in

$$X_T$$

.[20] WHITTLE (1963) shows that the MMSE estimator of

$$\hat{s}_t$$

is:

$$\hat{s}_t = k_s\frac{\psi_s(B)\psi_s(F)}{\psi(B)\psi(F)}x_t$$

, [22]

where

$$\psi(B) = \frac{\theta(B)}{\phi(B)}$$

,

$$F = B^{-1}$$

[20]MARAVALL, A. (2008c).

and

$$k_s = \frac{V(a_s)}{V(a)}$$

,

$$V(a_s)$$

is the variance of

$$a_{st}$$

and

$$V(a)$$

is the variance of

$$a_t$$

.

Expressing the

$$\psi(B)$$

polynomials as functions of the AR and MA polynomials, after cancelation of roots, the estimator of

$$s_t$$

can be expressed as:

$$\hat{s}_t = k_s \frac{\theta_s(B)\theta_s(F)\varphi_n(B)\delta_n(B)\varphi_n(F)\delta_n(F)}{\theta(B)\theta(F)} x_t$$

, [23]

where:

$$\nu_s(B,F) = k_s \frac{\theta_s(B)\theta_s(F)\varphi_n(B)\delta_n(B)\varphi_n(F)\delta_n(F)}{\theta(B)\theta(F)}$$

[24]

is a WK filter.

Equation [24] shows that the WK filter is two-sided (uses observations both from the past and from the future), centered (the number of points in the past is the same as in the future) and symmetric (for any k the weight applied to x_{t-k} and x_{t+k} is the same), which allows the phase effect to be avoided. Due to invertibility of $\theta(B)$ (and $\theta(F)$) the filter is convergent in the past and in the future.

The estimator can be presented as

$$\hat{s}_t = \nu_i\left(B, F\right) x_t$$

, [25]

where

$$\nu_i\left(B, F\right) = \nu_0 + \sum_{j=1}^{\infty} \nu_{ij}(B^j + F^j)$$

is the WK filter.

The example of the WK filters obtained for the pseudo-spectra of the series illustrated above is shown on the figure below: WK filters for components.

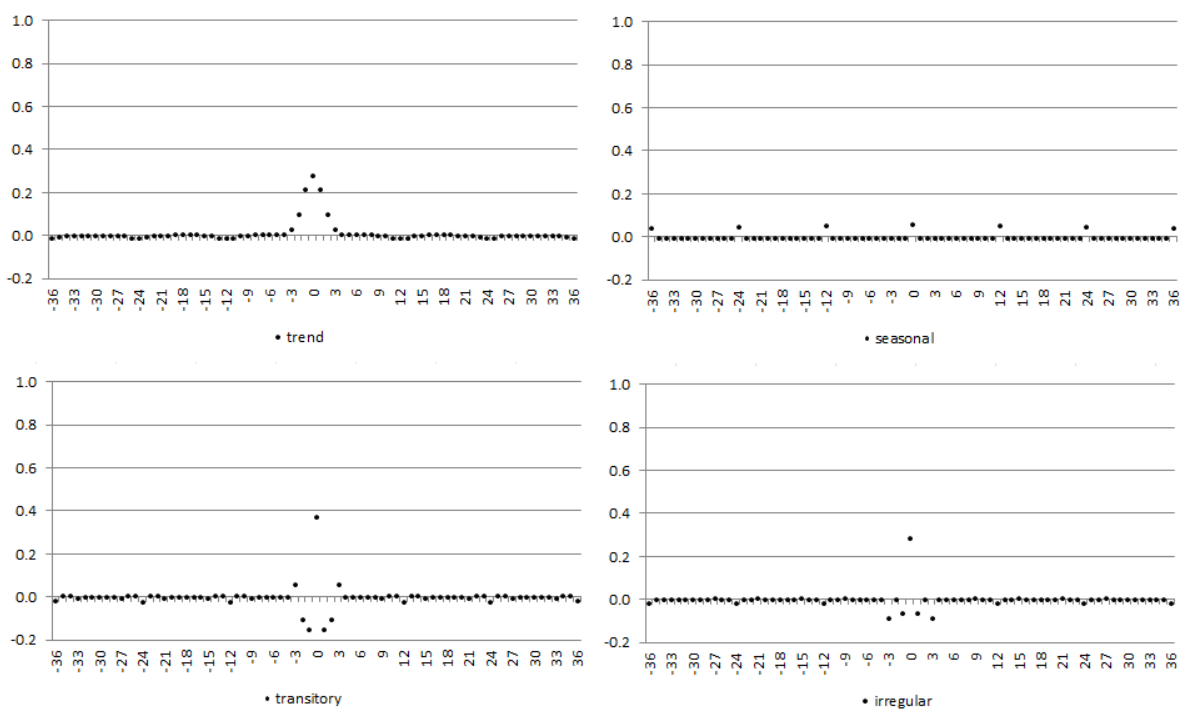

Figure 18.6: Text

WK filters for components

The WK filter from [24] can also be expressed as a ratio of two pseudo-autocovariance generating functions (p-ACGF). The p-ACGF function summarizes the sequence of absolutely summable autocovariances of a stationary process x_t (see section section Spectral Analysis).

The ACGF function of an ARIMA process is expressed as:

$$acgf(B) = \frac{\theta\left(B\right)\theta\left(F\right)}{\phi\left(B\right)\delta\left(B\right)\phi\left(F\right)\delta\left(F\right)}V(a)$$

[26]

And, the WK filter can be rewritten as:

$$\nu_s\left(B,F\right) = \frac{\gamma_s(B,F)}{\gamma(B,F)}$$

, [27]

where:

$$\gamma_s\left(B,F\right) = \frac{\theta_s\left(B\right)\theta_s\left(F\right)}{\phi_s\left(B\right)\delta_s\left(B\right)\phi_s\left(F\right)\delta_s\left(F\right)}V(a_s)$$

is the p-ACGF of

$$s_t$$

;

$\gamma\left(B,F\right) = \frac{\theta(B)\theta(F)}{\phi(B)\delta(B)\phi(F)\delta(F)}V(a)$ is the p-ACGF of x_t.

From [24] it can be seen that the WK filter depends on both the component and the series models. Consequently, the estimator of the component and the WK filter reflect the characteristic of data and by construction, the WK filter adapts itself to the series under consideration. Therefore, the ARIMA model is of particular importance for the SEATS method. Its misspecification results in an incorrect decomposition.

This adaptability, if the model has been correctly determined, avoids the dangers of under and overestimation with an ad-hoc filtering. For example, for the series with a highly stochastic seasonal component the filter adapts to the width of the seasonal peaks and the seasonally adjusted series does not display any spurious seasonality[21]. Examples of WK filters for stochastic and stable seasonal components are presented on the figure below.

WK filters for stable and stochastic seasonal components

The derivation of the components requires an infinite realization of x_t in the direction of the past and of the future. However, the convergence of the WK filter guarantees that, in practice, it could be approximated by a truncated (finite) filter and, in most applications, for large

$$k$$

the estimator for the central periods of the series can be safely seen as generated by the WK filter[22]:

[21] MARAVALL, A. (1995).
[22] MARAVALL, A., and PLANAS, C. (1999).

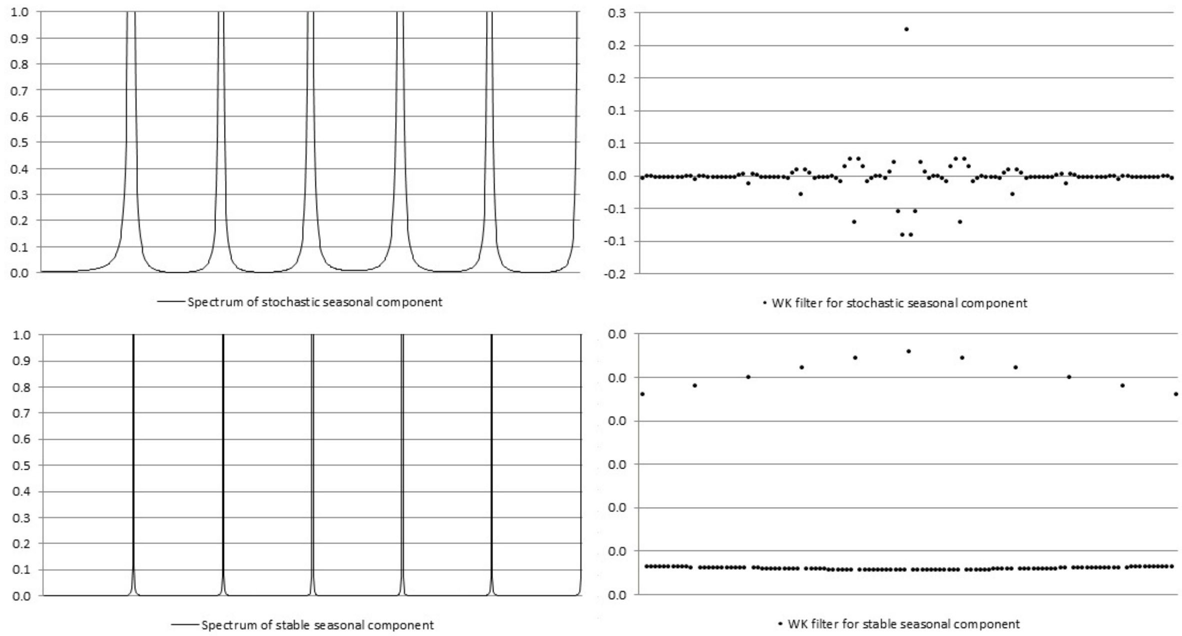

Figure 18.7: Text

$$\hat{s}_t = \nu_k x_{t-k} + ... + \nu_0 x_t + ... + \nu_k x_{t+k}$$

. [28]

When $T > 2L + 1$, where T is the last observed period, and L is an a priori number that typically expands between 3 and 5 years, the estimator expressed by [23] can be assumed as the final (historical) estimator for the central observations of the series[23]. In practice, the Wiener-Kolmogorov filter is applied to x_t extended with forecasts and backcasts from the ARIMA model. The final or historical estimator of

$$\hat{s}_t$$

, is obtained with a doubly infinite filter, and therefore contains an error

$$e_{st}$$

called final estimation error, which is equal

$$e_{st} = s_t - \hat{s}_t$$

.

[23]MARAVALL, A. (1998).

In the frequency domain, the Wiener-Kolmogorov filter $\nu(B, F)$ that provides the final estimator of s_t is expressed as the ratio of the s_t and x_t pseudo-spectra:

$$\tilde{\nu}\left(\omega\right)=\frac{g_s(\omega)}{g_x(\omega)}$$

. [29]

The function $\tilde{\nu}\left(\omega\right)$ is also referred as the gain of the filter.[24] GÓMEZ, V., and MARAVALL, A. (2001a) show that when for some frequency the signal (the seasonally adjusted series) dominates the noise (seasonal fluctuations) the gain $\tilde{\nu}\left(\omega\right)$ approaches 1. On the contrary, when for some frequency the noise dominates the gain $\tilde{\nu}\left(\omega\right)$ approaches 0.

The spectrum of the estimator of the seasonal component is expressed as:

$$g_{\hat{s}}\left(\omega\right)=\left[\frac{g_s(\omega)}{g_x(\omega)}\right]^2 g_x(\omega)$$

, [30]

where $\left[\tilde{\nu}\left(\omega\right)\right]^2=\left[\frac{g_s(\omega)}{g_x(\omega)}\right]^2=\left[\frac{g_s(\omega)}{g_s(\omega)+g_n(\omega)}\right]^2=\left[\frac{1}{1+\frac{1}{r(\omega)}}\right]^2$ is the squared gain of the filter and $r\left(\omega\right)=\frac{g_s(\omega)}{g_n(\omega)}$ represents the signal-to-noise ratio.

For each ω, the MMSE estimation gives the signal-to-noise ratio. If this ratio is high, then the contribution of that frequency to the estimation of the signal will be also high. Assume that the trend is a signal that needs to be extracted from a seasonal time series. Then $R\left(0\right)=1$ and the frequency $\omega=0$ will only be used for trend estimations. For seasonal frequencies $R\left(\omega\right)=0$, so that these frequencies are ignored in computing the trend resulting in spectral zeros in $g_{\hat{s}}\left(\omega\right)$. For this reason, unlike the spectrum of the component, the component spectrum contains dips as it can be seen on the figure below: Component spectrum and estimator spectrum for trend.

Component spectrum and estimator spectrum for trend

From the equation [29] it is clear that the squared gain of the filter determines how the variance of the series contributes to the variance of the seasonal component for the different frequencies. When $\tilde{\nu}\left(\omega\right)=1$, the full variation of x_t for that frequency is passed to

$$\hat{s}_t$$

, while if

$$\tilde{\nu}\left(\omega\right)=0$$

[24] GÓMEZ, V., and MARAVALL, A. (2001a).

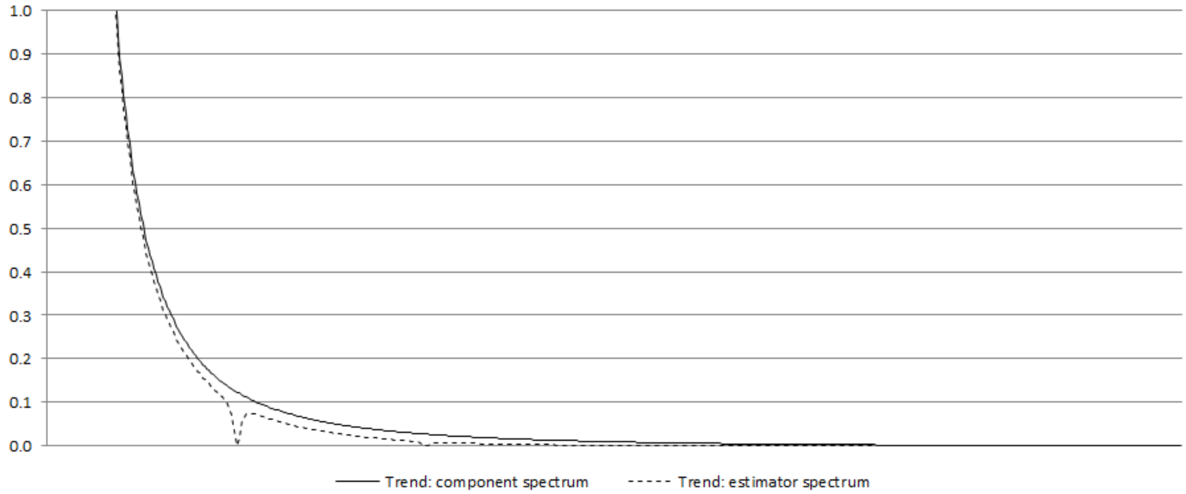

Figure 18.8: Text

the variation of x_t for that frequency is fully ignored in the computation of

$$\hat{s}_t$$

. These two cases are well illustrated by the figure below that shows the square gain of the WK filter for two series already analysed in the figure above (Figure: WK filters for stable and stochastic seasonal components).

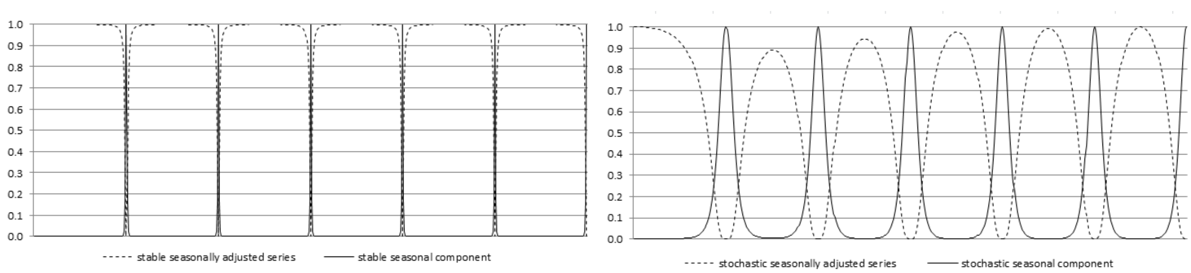

Figure 18.9: Text

The squared gain of the WK filter for stable and stochastic seasonal components.

Since $r(\omega) \geq 0$, then $\tilde{\nu}(\omega) \leq 1$ and from [29] it can be derived that $g_{\hat{s}}(\omega) = \tilde{\nu}(\omega) g_s(\omega)$. As a result, the estimator will always underestimate the component, i.e. it will be always more stable that the component.[25]

Since $g_{\widehat{n}}(\omega) < g_n(\omega)$ and $g_{\hat{s}}(\omega) < g_s(\omega)$ the expression: $g_x(\omega) - [g_{\widehat{n}}(\omega) + g_{\hat{s}}(\omega)] \geq 0$ is the cross-spectrum. As it is positive, the MMSE yields correlated estimators. This effect emerges

[25]Ibid.

since variance of estimator is smaller than the variance of component. Nevertheless, if at least one non-stationary component exists, cross-correlations estimated by TRAMO-SEATS will tend to zero as cross-covariances between estimators of the components are finite. In practice, the inconvenience caused by this property will likely be of little relevance.

Preliminary estimators for the components

GÓMEZ, V., and MARAVALL, A. (2001a) point out that *the properties of the estimators have been derived for the final (or historical) estimators. For a finite (long enough) realization, they can be assumed to characterize the estimators for the central observations of the series, but for periods close to the beginning of the end the filter cannot be completed and some preliminary estimator has to be used.* Indeed, the historical estimator shown in [28] is obtained for the central periods of the series. However, when t approaches T (last observation), the WK filter requires observations, which are not available yet. For this reason a preliminary estimator needs to be used.

To introduce preliminary estimators let us consider a semi-finite realization $[x_{-\infty},\dots x_T]$, where T is the last observed period. The preliminary estimator of

$$x_{\mathrm{it}}$$

obtained at T

$$(T-t=k\geq 0)$$

can be expressed as

$$\hat{x}_{it|t+k}=\nu_i\left(B,F\right)x^e_{t|T}$$

, [31]

where

$$\nu_i\left(B,F\right)$$

is the WK filter and

$$x^e_{t|T}$$

is the extended series, such that $x^e_{t|T}=x_t$ for $t\leq T$ and

$$x^e_{t|T}=\hat{x}_{t|T}$$

for

$$t>T$$

, where

$$\hat{x}_{t|T}$$

denotes the forecast of x_t obtained at period T.

The future k values necessary to apply the filter are not yet available and are replaced by their optimal forecasts from the ARIMA model on

$$x_t$$

. When

$$k = 0$$

the preliminary estimator becomes the concurrent estimator. As the forecasts are linear functions of present and past observations of

$$x_t$$

, the preliminary estimator

$$\hat{x}_{it}$$

will be a truncated asymmetric filter applied to

$$x_t$$

that generates a phase effect[26].

When a new observation

$$x_{T+1}$$

becomes available the forecast

$$\hat{x}_{T+1|T}$$

is replaced by the observation and the forecast

$$\hat{x}_{iT+j|T}$$

,

$$j > 1$$

are updated to

$$x_{T+j|T+1}$$

resulting in the revision error[27]. The total error in the preliminary estimator

$$d_{it|t+k}$$

is expressed as a sum of the final estimation error (

$$e_{it}$$

) and the revision error (

$$r_{it|t+k}$$

[26]KAISER, R., and MARAVALL, A. (2000).
[27]MARAVALL, A. (1995).

), i.e.:

$$d_{it|t+k} = x_{it} - \hat{x}_{it|t+k} = (x_{it} - \hat{x}_{it}) + \left(\hat{x}_{it} - \hat{x}_{it|t+k}\right) = e_{it} + r_{it|t+k}$$

, [32]

where:

$$x_{it} - i^{th}$$

component;

$$\hat{x}_{it|t+k}$$

- the estimator of

$$x_{it}$$

when the last observation is

$$x_{t+k}$$

.

Therefore the preliminary estimator is subject not only to the final error but also to a revision error, which are orthogonal to each other[28]. The revision error decreases as

$$k$$

increases, until it can be assumed equal to 0 for large enough

$$k$$

.

It's worth remembering that SEATS estimates the unobservable components of the time series so the "true" components are never observed. Therefore, MARAVALL, A. (2009) stresses that *the error in the historical estimator is more of academic rather than practical interest. In practice, interest centres on revisions. (...) the revision standard deviation will be an indicator of how far we can expect to be from the optimal estimator that will be eventually attained, and the speed of convergence of* $\theta(B)^{-1}$ *will dictate the speed of convergence of the preliminary estimator to the historical one.* The analysis of an error is therefore useful for making decision concerning the revision policy, including the policy for revisions and horizon of revisions.

[28] MARAVALL, A. (2009).

18.5 PsiE-weights

The estimator of the component is calculated as

$$\hat{x}_{it} = \nu_s (B, F) \, x_t$$

. By replacing

$$x_{it} = \frac{\theta(B)}{\gamma(B)\delta(B)} a_t$$

, the component estimator can be expressed as[29]:

$$\hat{x}_{it} = \xi_s (B, F) \, a_t$$

, [33]

where $\xi_s (B, F) = ... + \xi_j B^j + ... + \xi_1 B + \xi_0 + \xi_{-1} F \ldots \xi_{-j} F^j + ...$

This representation shows the estimator as a filter applied to the innovation

$$a_t$$

, rather than on the series

$$x_t$$

[30]. Hence, the filter from [32] can be divided into two components: the first one, i.e.

$$... + \xi_j B^j + ... + \xi_1 B + \xi_0$$

, applies to prior and concurrent innovations, the second one, i.e.

$$\xi_{-1} F + ... + \xi_{-j} F^j$$

applies to future (i.e. posterior to

$$t$$

) innovations. Consequently,

$$\xi_j$$

determines the contribution of

$$a_{t-j}$$

to

$$\hat{s}_t$$

while

$$\xi_{-j}$$

[29]The section is based on KAISER, R., and MARAVALL, A. (2000).
[30]See section PsiE-weights. For further details see MARAVALL, A. (2008).

determines the contribution of

$$a_{t+j}$$

to

$$\hat{s}_t$$

. Finally, the estimator of the component can be expressed as:

$$\hat{x}_{it} = \xi_i(B)^{-}a_t + \xi_i(F)^{+}a_{t+1}$$

, [34]

where:

$\xi_i(B)^{-}a_t$ is an effect of starting conditions, present and past innovations in series;

$\xi_i(F)^{+}a_{t+1}$ is an effect of future innovations.

For the two cases already presented in figure *WK filters for stable and stochastic seasonal components* and figure *The squared gain of the WK filter for stable and stochastic seasonal components* above, the psi-weights are shown in the figure below.

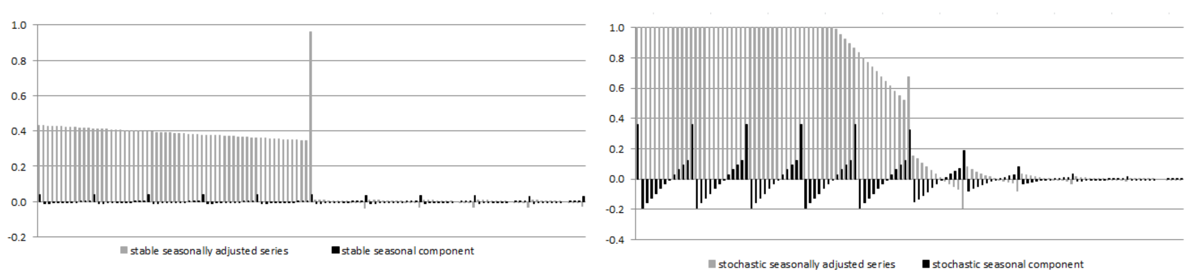

Figure 18.10: Text

It can be shown that

$$\xi_{-1}, \dots, \xi_{-j}$$

are convergent and

$$\xi_j, \dots, \xi_1, \xi_0$$

are divergent. From [33] , the concurrent estimator is equal to

$$\hat{x}_{it|t} = E_t x_{it} = E_t \hat{x}_{it} = \xi_i(B)^{-}a_t$$

, [35]

so that the revision

$$r_{it} = \hat{x}_{it} - \hat{x}_{it|t} = \xi_i(F)^{+}a_{t+1}$$

[36]

is a zero-mean stationary MA process. As a result, historical and preliminary estimators are cointegrated. From expression [25] the relative size of the full revision and the speed of convergence can be obtained.

19 Local Polynomials Methods for Trend Estimation

19.1 Chapter building process

This chapter provides details on the the statistical methods whereas practical estimation steps are described "here" (Chap On trend estimation) and are run with the rjfilters or rjd3highfreq packages.

20 Tests

20.1 Introduction

This chapter describes all the tests available in JDemetra+, via Graphical User interface and/or R packages. An outline of the underlying theoretical principles of each test is provided.

The procedure to apply these tests in context is described when their use is relevant in the chapters dedicated to algorithms description, mainly on seasonal adjustment.(not sure) Links to use in context are available below.

20.2 Tests on residuals

table with all tests by purpose and accessibility

Table 20.1: Tests on Residuals

Test	Purpose	GUI	R package
Ljung-Box	autocorrelation		
Box-Pierce	autocorrelation		
Doornik-Hansen	normality		

20.2.1 Ljung-Box

The Ljung-Box Q-statistics are given by:

$$\mathrm{LB}\left(k\right) = n \times (n+2) \times \sum_{k=1}^{K} \frac{\rho_{a,k}^{2}}{n-k}$$

, [1]

where $\rho_{a,k}^{2}$ is the autocorrelation coefficient at lag k of the residuals $\hat{a}_t$, n is the number of terms in differenced (? differenciated ?) series, K is the maximum lag being considered, set in JDemetra+ to 24 (monthly series) or 8 (quarterly series).

If the residuals are random (which is the case for residuals from a well specified model), they will be distributed as $\chi^2_{(K-m)}$, where m is the number of parameters in the model which has been fitted to the data. (edit: not the residuals, but $\hat{\rho}$)

The Ljung-Box and Box-Pierce tests sometimes fail to reject a poorly fitting model. Therefore, care should be taken not to accept a model on a basis of their results. For the description of autocorrelation concept see section Autocorrelation function and partial autocorrelation function.

20.2.2 Box-Pierce

The Box-Pierce Q-statistics are given by:

$$\mathrm{BP}\,(k) = n \sum_{k=1}^{K} \rho^2_{a,k}$$

where:

$\rho^2_{a,k}$ is the autocorrelation coefficient at lag k of the residuals $\hat{a}_t$.

n is the number of terms in differenced (differenciated?) series;

K is the maximum lag being considered, set in JDemetra+ to 24 (monthly series) or 8 (quarterly series).

If the residuals are random (which is the case for residuals from a well specified model), they will be distributed as $\chi^2_{(K-m)}$ degrees of freedom, where m is the number of parameters in the model which has been fitted to the data.(edit: same as above)

20.2.3 Dornik-Hansen

The Doornik-Hansen test for multivariate normality (DOORNIK, J.A., and HANSEN, H. (2008)) is based on the skewness and kurtosis of multivariate data that is transformed to ensure independence. It is more powerful than the Shapiro-Wilk test for most tested multivariate distributions[1].

The skewness and kurtosis are defined, respectively, as:

$$s = \frac{m_3}{\sqrt{m_2}^3}$$

and $k = \frac{m_{4}}{m_{2}^{2}},\ $where: $m_i = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^i$ $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and n is a number of (non-missing) residuals.

[1] The description of the test derives from DOORNIK, J.A., and HANSEN, H. (2008).

The Doornik-Hansen test statistic derives from SHENTON, L.R., and BOWMAN, K.O. (1977) and uses transformed versions of skewness and kurtosis.

The transformation for the skewness s into z_1 is as in D'AGOSTINO, R.B. (1970):

$$\beta = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}$$

$$\omega^2 = -1 + \sqrt{2(\beta-1)}$$

$$\delta = \frac{1}{\sqrt{\log\left(\omega^2\right)}}$$

$$y = s\sqrt{\frac{(\omega^2-1)(n+1)(n+3)}{12(n-2)}}$$

$$z_1 = \delta log(y + \sqrt{y^2-1})$$

The kurtosis k is transformed from a gamma distribution to χ^2, which is then transformed into standard normal z_2 using the Wilson-Hilferty cubed root transformation:

$$\delta = (n-3)(n+1)(n^2+15n-4)$$

$$a = \frac{(n-2)(n+5)(n+7)(n^2+27n-70)}{6\delta}$$

$$c = \frac{(n-7)(n+5)(n+7)(n^2+2n-5)}{6\delta}$$

$$l = \frac{(n+5)(n+7)(n^3+37n^2+11n-313)}{12\delta}$$

$$\alpha = a + c \times s^2$$

$$\chi = 2l(k-1-s^2)$$

$$z_2 = \sqrt{9\alpha}\left(\frac{1}{9\alpha} - 1 + \sqrt[3]{\frac{\chi}{2\alpha}}\right)$$

Finally, the Doornik-Hansen test statistic is defined as the sum of squared transformations of the skewness and kurtosis. Approximately, the test statistic follows a χ^2distribution, i.e.:

$$DH = z_1^2 + z_2^2 \sim \chi^2(2)$$

20.3 Seasonality tests

table with all tests by purpose and accessibility

Table 20.2: Seasonality tests

Test	Purpose	GUI	R package
QS test	Autocorrelation at seasonal lags		
F-test with seasonal dummies	Stable seasonality		
Identification of spectral peaks	Seasonal frequencies		
Friedman test	Stable seasonality		
Two-way variance analysis	Moving seasonality		

20.3.0.1 QS Test on autocorrelation at seasonal lags

The QS test is a variant of the Ljung-Box test computed on seasonal lags, where we only consider positive auto-correlations

More exactly,

$$QS = n\left(n+2\right)\sum_{i=1}^{k}\frac{\left[\max\left(0, \hat{\gamma}_{i\cdot l}\right)\right]^2}{n - i \cdot l}$$

where

$$k = 2$$

, so only the first and second seasonal lags are considered. Thus, the test would checks the correlation between the actual observation and the observations lagged by one and two years. Note that

$$l = 12$$

when dealing with monthly observations, so we consider the autocovariances

$$\hat{\gamma}_{12}$$

and

$$\hat{\gamma}_{24}$$

alone. In turn,

$$k = 4$$

in the case of quarterly data.

Under H0, which states that the data are independently distributed, the statistics follows a

$$\chi\left(k\right)$$

distribution. However, the elimination of negative correlations makes it a bad approximation. The p-values would be given by $P(\chi^2\left(k\right) > Q)$ for $k = 2$. As $P(\chi^{2}(2)) > 0.05 = 5.99146$ and $P(\chi^{2}(2)) > 0.01 = 9.21034$, $QS > 5.99146$ and $QS > 9.21034$ would suggest rejecting the null hypothesis at 95% and 99% significance levels, respectively.

20.3.0.2 Modification

Maravall (2012) proposes approximate the correct distribution (p-values) of the QS statistic using simulation techniques. Using 1000K replications of sample size 240, the correct critical values would be 3.83 and 7.09 with confidence levels of 95% and 99%, respectively (lower than the 5.99146 and 9.21034 shown above). For each of the simulated series, he obtains the distribution by assuming $QS = 0$ when

$$\hat{\gamma}_{12}$$

, so in practice this test will detect seasonality only when any of these conditions hold: - Statistically significant positive autocorrelation at lag 12 - Nonnegative sample autocorrelation at lag 12 and statistically significant positive autocorrelation at lag 24

20.3.0.3 Use

The test can be applied directly to any series by selecting the option *Statistical Methods » Seasonal Adjustment » Tools » Seasonality Tests*. This is an example of how results are displayed for the case of a monthly series:

The test can be applied to the input series before any seasonal adjustment method has been applied. It can also be applied to the seasonally adjusted series or to the irregular component.

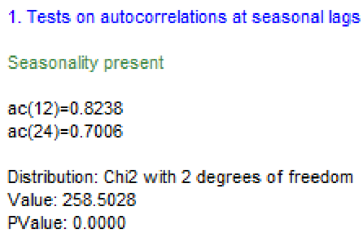

Figure 20.1: qs

20.3.0.4 References

- LJUNG G. M. and G. E. P. BOX (1978). "On a Measure of a Lack of Fit in Time Series Models". Biometrika 65 (2): 297–303. doi:10.1093/biomet/65.2.297
- MARAVALL, A. (2011). "Seasonality Tests and Automatic Model Identification in Tramo-Seats". Manuscript
- MARAVALL, A. (2012). "Update of Seasonality Tests and Automatic Model Identification in TRAMO-SEATS". Bank of Spain (November 2012)

20.3.1 F-test on seasonal dummies

The F-test on seasonal dummies checks for the presence of deterministic seasonality. The model used here uses seasonal dummies (mean effect and 11 seasonal dummies for monthly data, mean effect and 3 for quarterly data) to describe the (possibly transformed) time series behaviour. The test statistic checks if the seasonal dummies are jointly statistically not significant. When this hypothesis is rejected, it is assumed that the deterministic seasonality is present and the test results are displayed in green.

This test refers to Model-Based $\chi^{2}\ $and F-tests for Fixed Seasonal Effects proposed by LYTRAS, D.P., FELDPAUSCH, R.M., and BELL, W.R. (2007) that is based on the estimates of the regression dummy variables and the corresponding t-statistics of the RegARIMA model, in which the ARIMA part of the model has a form (0,1,1)(0,0,0). The consequences of a misspecification of a model are discussed in LYTRAS, D.P., FELDPAUSCH, R.M., and BELL, W.R. (2007).

For a monthly time series the RegARIMA model structure is as follows:

$$(1-B)\left(y_t-\beta_1 M_{1,t}-...-\beta_{11} M_{11,t}-\gamma X_t\right)=\mu+(1-B)a_t$$

, [1]

where:

$$M_{j,t} = \begin{cases} 1 & \text{in month } j = 1, \dots, 11 \\ -1 & \text{in December} \\ 0 & \text{otherwise} \end{cases} \quad \text{- dummy variables;}$$

y_t – the original time series;

B – a backshift operator;

X_t – other regression variables used in the model (e.g. outliers, calendar effects, user-defined regression variables, intervention variables);

μ – a mean effect;

a_t – a white-noise variable with mean zero and a constant variance.

In the case of a quarterly series the estimated model has a form:

$$(1 - B)\left(y_t - \beta_1 M_{1,t} - \dots - \beta_3 M_{3,t} - \gamma X_t\right) = \mu + (1 - B)a_t$$

, [2]

where:

$$M_{j,t} = \begin{cases} 1 & \text{in quarter} j = 1, \dots, 3 \\ -1 & \text{in the fourth quarter} \\ 0 & \text{otherwise} \end{cases} \text{ - dummy variables;}$$

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 $\hat{\chi}^2 = \hat{\beta}'[Var(\hat{\beta})\ )^{-1}]\hat{\beta}$ in this case compared to critical values from a χ^2(df)-distribution, with degrees of freedom \$df = 11\ \$(monthly series) or $df = 3$ (quarterly series). Since the $Var(\hat{\beta})$ computed using the estimated variance of α_t may be very different from the actual variance in small samples, this test is corrected using the proposed F statistic:

$$F = \frac{\hat{\chi}^2}{s - 1} \times \frac{n - d - k}{n - d}$$

where n is the sample size, d is the degree of differencing, s is time series frequency (12 for a monthly series, 4 for a quarterly series) and k is the total number of regressors in the RegARIMA model (including the seasonal dummies $\mathrm{M}_{j,t}$ and the intercept).

This statistic follows a

$$F_{s-1,n-d-k}$$

distribution under the null hypothesis.

20.3.2 Identification of spectral peaks

link to relevant part in spectral analysis chapter ?

20.3.3 Friedman test for stable seasonality test

The Friedman test is a non-parametric method for testing that samples are drawn from the same population or from populations with equal medians. The significance of the month (or quarter) effect is tested. The Friedman test requires no distributional assumptions. It uses the rankings of the observations. If the null hypothesis of no stable seasonality is rejected at the 0.10% significance level then the series is considered to be seasonal and the test's outcome is displayed in green.

The test statistic is constructed as follows. Consider first the matrix of data

$$\{x_{ij}\}_{n\times k}$$

with

$$n$$

rows (the blocks, i.e. number of years in the sample),

$$k$$

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

$$\{r_{ij}\}_{n\times k}$$

, where the entry

$$r_{ij}$$

is the rank of

$$x_{ij}$$

within block

$$i$$

.

The test statistic is given by

$$Q=\frac{SS_t}{SS_e}$$

where

$$SS_t = n \sum_{j=1}^{k} (\bar{r}_{.j} - \bar{r})^2$$

and

$$SS_e = \frac{1}{n(k-1)} \sum_{i=1}^{n} \sum_{j=1}^{k} (r_{ij} - \bar{r})^2$$

It represents the variance of the average ranking across treatments j relative to the total.

Under the hypothesis of no seasonality, all months can be equally treated. For the sake of completeness: -

$$\bar{r}_{.j}$$

is the average ranks of each treatment (month) j within each block (year) - The average rank is given by

$$\bar{r} = \frac{1}{nk} \sum_{i=1}^{n} \sum_{j=1}^{k} (r_{ij})$$

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

$$P(\chi^2_{k-1} > Q)$$

.

20.3.3.1 Use

The test can be applied directly to any series by selecting the option *Statistical Methods » Seasonal Adjustment » Tools » Seasonality Tests*. This is an example of how results are displayed for the case of a monthly series:

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 outcome of the test is displayed in green.

The test can be applied to the input series before any seasonal adjustment method has been applied. It can also be applied to the seasonally adjusted series or to the irreguar component.

Figure 20.2: friedman

In the case of X-13ARIMA-SEATS, the test is applied to the preliminary estimate of the unmodified Seasonal-Irregular component[2] (time series shown in Table B3). In this estimate, the number of observations is lower than in the final estimate of the unmodified Seasonal-Irregular component. Thus, the number of degrees of freedom in the stable seasonality test is lower than the number of degrees of freedom in the test for the presence of seasonality assuming stability. For example, X-13ARIMA-SEATS uses a centred moving average of order 12 to calculate the preliminary estimation of trend. Consequently, the first six and last six points in the series are not computed at this stage of calculation. The preliminary estimation of the trend is then used for the calculation of the preliminary estimation of the unmodified Seasonal-Irregular.

20.3.3.2 Related tests

- When using this kind of design for a binary response, one instead uses the Cochran's Q test.
- Kendall's W is a normalization of the Friedman statistic between 0 and 1.
- The Wilcoxon signed-rank test is a nonparametric test of non-independent data from only two groups.

20.3.3.3 References

- Friedman, Milton (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:10.2307/2279372. JSTOR 2279372.

- Friedman, Milton (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:10.2307/2279169. JSTOR 2279169.

[2]The unmodified Seasonal-Irregular component corresponds to the Seasonal-Irregular factors with the extreme values.

- Friedman, Milton (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:10.1214/aoms/1177731944. JSTOR 2235971.

20.3.4 Moving seasonality test

The evolutive seasonality test is based on a two-way analysis of variance model. The model uses the values from complete years only. Depending on the decomposition type for the Seasonal – Irregular component it uses [1] (in the case of a multiplicative model) or [2] (in the case of an additive model):

$$\left|\mathrm{SI}_{\mathrm{ij}} - 1\right| = X_{\mathrm{ij}} = b_i + m_j + e_{\mathrm{ij}}$$

, [1]

$$\left|\mathrm{SI}_{\mathrm{ij}}\right| = X_{\mathrm{ij}} = b_i + m_j + e_{\mathrm{ij}}$$

, [2]

where:

m_j – the monthly or quarterly effect for j-th period, $j = (1, \ldots, k)$, where $k = 12$ for a monthly series and $k = 4$ for a quarterly series;

b_j – the annual effect i, $(i = 1, \ldots, N)$ where N is the number of complete years;

e_{ij} – the residual effect.

The test is based on the following decomposition:

$$S^2 = S_A^2 + S_B^2 + S_R^2,$$

[3]

where:

$$S^2 = \sum_{j=1}^{k} \sum_{i=1}^{N} \left(\overline{X}_{\mathrm{ij}} - \overline{X}_{\bullet\bullet}\right)^2$$

–the total sum of squares;

$$S_A^2 = N \sum_{j=1}^{k} \left(\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet}\right)^2$$

– the inter-month (inter-quarter, respectively) sum of squares, which mainly measures the magnitude of the seasonality;

$$S_B^2 = k\sum_{i=1}^{N}\left(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet}\right)^2$$

– the inter-year sum of squares, which mainly measures the year-to-year movement of seasonality;

$$S_R^2 = \sum_{i=1}^{N}\sum_{j=1}^{k}\left(\overline{X}_{\mathrm{ij}} - \overline{X}_{i\bullet} - \overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet}\right)^2$$

– the residual sum of squares.

The null hypothesis \$H_{0}\ \$is that $b_1 = b_2 = ... = b_N$ which means that there is no change in seasonality over the years. This hypothesis is verified by the following test statistic:

$$F_M = \frac{\frac{S_B^2}{(n-1)}}{\frac{S_R^2}{(n-1)(k-1)}}$$

which follows an F-distribution with $k-1$ and $n-k$ degrees of freedom.

20.3.5 Combined seasonality test

This test combines the Kruskal-Wallis test along with test for the presence of seasonality assuming stability (F_S), and evaluative seasonality test for detecting the presence of identifiable seasonality (F_M). Those three tests are calculated using the final unmodified SI component. The main purpose of the combined seasonality test is to check whether the seasonality of the series is identifiable. For example, the identification of the seasonal pattern is problematic if the process is dominated by highly moving seasonality[3]. The testing procedure is shown in the figure below.

Combined seasonality test, source: LADIRAY, D., QUENNEVILLE, B. (2001)

[3]DAGUM, E.B. (1987).

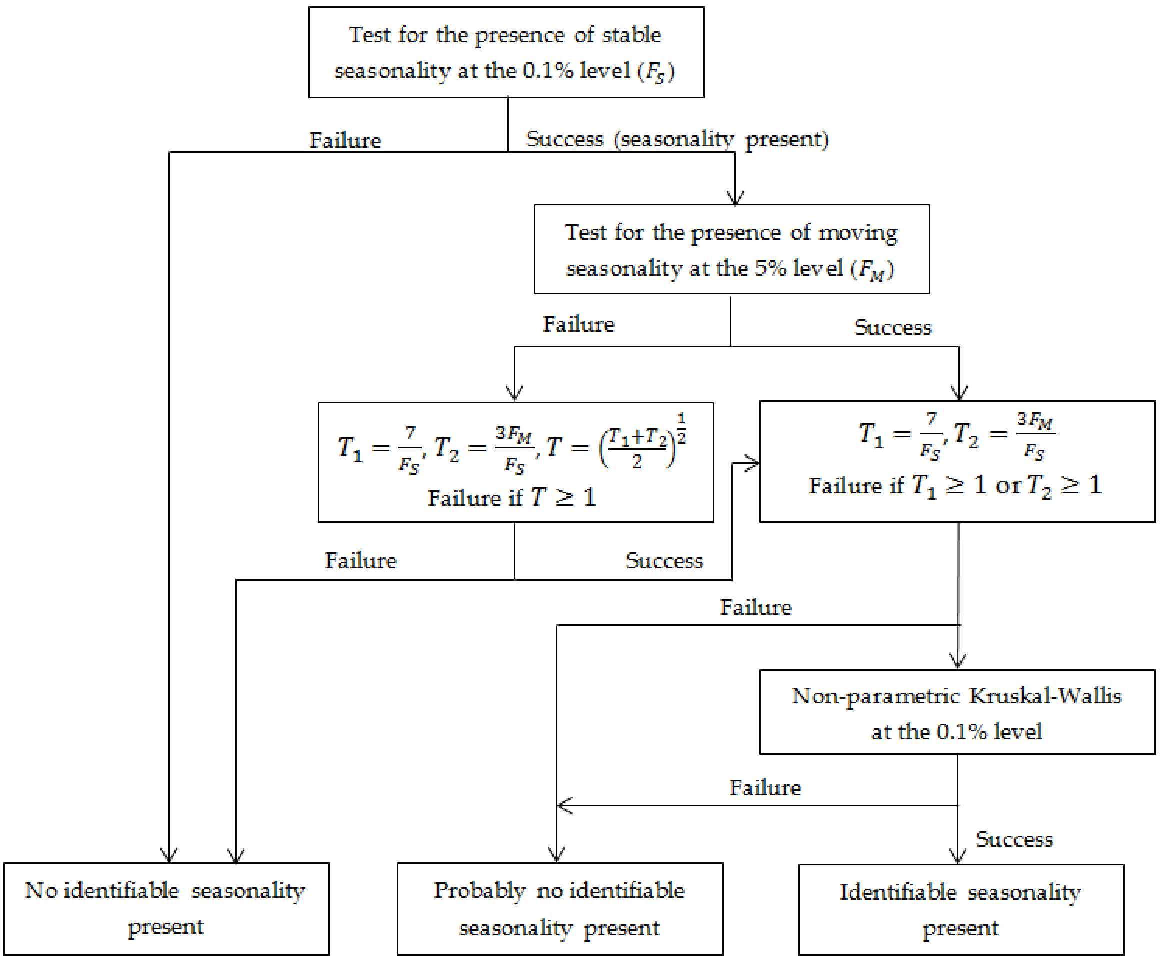

Figure 20.3: Text

21 State space modelling

21.1 Introduction

State space forms play a central role in JD+. In a first point, we present the model used throughout the library. In a second point, we give an overview of the current implementation. The specificity of the state space framework of JD+ is stressed and the main modules are shortly presented. We finish the document with two examples based on the state space framework of JD+. The first one explains the basic structural models implemented in the "nbdemetra-sa-advanced" plug-in of JD+. The second one shows how the framework can be used to implement time-varying trading days models.

21.2 State space forms (SSF)

21.2.1 General form

The general linear gaussian state-space model can be written in many different ways. The measurement equation and the state equation considered in JD+ 3.0 are presented below.

$$\begin{aligned} y_t &= Z_t\alpha_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2 H_t) \\ \alpha_{t+1} &= T_t\alpha_t + \mu_t, \quad \mu_t \sim \mathcal{N}(0, \sigma^2 V_t) \end{aligned}$$

y_t is the observation at period t, α_t is the state vector. ϵ_t, μ_t are assumed to be serially independent at all time points and independent between them at all time points.

The residuals of the state equation will be modelled as $\mu_t = S_t\xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma^2 I)$.

In other words, $V_t = S_t S_t'$.

The initial conditions of the filter are defined as follows:

$$\alpha_{-1} = a_{-1} + A_{-1}\delta + \mu_{-1}$$

$$\delta \sim \mathcal{N}(0, \sigma^2\kappa I)$$

$$\mu_{-1} \sim \mathcal{N}(0, \sigma^2 P_*)$$

where κ is arbitrary large.

P_* is the variance of the stationary part of the initial state vector and models the diffuse part.

The model used in JD+ is quasi-identical to the model discussed in Durbin-Koopman.

22 Methods for Temporal disaggregation and benchmarking

22.1 Benchmarking Underlying Theory

Benchmarking[1] is a procedure widely used when for the same target variable the two or more sources of data with different frequency are available. Generally, the two sources of data rarely agree, as an aggregate of higher-frequency measurements is not necessarily equal to the less-aggregated measurement. Moreover, the sources of data may have different reliability. Usually it is thought that less frequent data are more trustworthy as they are based on larger samples and compiled more precisely. The more reliable measurement is considered as a benchmark.

Benchmarking also occurs in the context of seasonal adjustment. Seasonal adjustment causes discrepancies between the annual totals of the seasonally unadjusted (raw) and the corresponding annual totals of the seasonally adjusted series. Therefore, seasonally adjusted series are benchmarked to the annual totals of the raw time series[2]. Therefore, in such a case benchmarking means the procedure that ensures the consistency over the year between adjusted and non-seasonally adjusted data. It should be noted that the '*ESS Guidelines on Seasonal Adjustment*' (2015) do not recommend benchmarking as it introduces a bias in the seasonally adjusted data. Also the U.S. Census Bureau points out that: *Forcing the seasonal adjustment totals to be the same as the original series annual totals can degrade the quality of the seasonal adjustment, especially when the seasonal pattern is undergoing change. It is not natural if trading day adjustment is performed because the aggregate trading day effect over a year is variable and moderately different from zero.*[3] Nevertheless, some users may prefer the annual totals for the seasonally adjusted series to match the annual totals for the original, non-seasonally adjusted series[4]. According to the '*ESS Guidelines on Seasonal Adjustment*' (2015), the only benefit of this approach is that there is consistency over the year between adjusted and non-seasonally adjusted data; this can be of particular interest when low-frequency (e.g. annual) benchmarking figures officially exist (e.g. National Accounts, Balance of Payments, External Trade, etc.) where user needs for time consistency are stronger.

[1]Description of the idea of benchmarking is based on DAGUM, B.E., and CHOLETTE, P.A. (1994) and QUENNEVILLE, B. et all (2003). Detailed information can be found in: DAGUM, B.E., and CHOLETTE, P.A. (2006).

[2]DAGUM, B.E., and CHOLETTE, P.A. (2006).

[3]'*X-12-ARIMA Reference Manual'* (2011).

[4]HOOD, C.C.H. (2005).

The benchmarking procedure in JDemetra+ is available for a single seasonally adjusted series and for an indirect seasonal adjustment of an aggregated series. In the first case, univariate benchmarking ensures consistency between the raw and seasonally adjusted series. In the second case, the multivariate benchmarking aims for consistency between the seasonally adjusted aggregate and its seasonally adjusted components.

Given a set of initial time series

$$\left\{z_{i,t}\right\}_{i\in I}$$

, the aim of the benchmarking procedure is to find the corresponding

$$\left\{x_{i,t}\right\}_{i\in I}$$

that respect temporal aggregation constraints, represented by

$$X_{i,T} = \sum_{t\in T} x_{i,t}$$

and contemporaneous constraints given by

$$q_{k,t} = \sum_{j\in J_k} w_{\mathrm{kj}} x_{j,t}$$

or, in matrix form:

$$q_{k,t} = w_k x_t$$

.

The underlying benchmarking method implemented in JDemetra+ is an extension of Cholette's[5] method, which generalises, amongst others, the additive and the multiplicative Denton procedure as well as simple proportional benchmarking.

The JDemetra+ solution uses the following routines that are described in DURBIN, J., and KOOPMAN, S.J. (2001):

- The multivariate model is handled through its univariate transformation,
- The smoothed states are computed by means of the disturbance smoother.

The performance of the resulting algorithm is highly dependent on the number of variables involved in the model ($\propto\ n^3$). The other components of the problem (number of constraints, frequency of the series, and length of the series) are much less important ($\propto\ n$).

From a theoretical point of view, it should be noted that this approach may handle any set of linear restrictions (equalities), endogenous (between variables) or exogenous (related to external values), provided that they don't contain incompatible equations. The restrictions can also be relaxed for any period by considering their "observation" as missing. However, in

[5] CHOLETTE, P.A. (1979).

practice, it appears that several kinds of contemporaneous constraints yield unstable results. This is more especially true for constraints that contain differences (which is the case for non-binding constraints). The use of a special square root initializer improves in a significant way the stability of the algorithm.

22.2 Temporal Disaggragation underlying Theory

References