JDemetra+ online documentation

Stace documentation group

12/01/2022

Table of contents

Preface

Welcome to the JDemetra+ online documentation.

JDemetra+ is a software for seasonal adjustment and other time series functions, developed in Eurostat's "Centre of Excellence on Statistical Methods and Tools".

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

1 JDemetra+ Software

ressources for descrption - R tools WP - desp in esp

1.1 A library of algorithms for time series related functions ? needs ?

You can learn more about the history of the project here (link to below)

1.2 Structure of this book

This book is divided in four parts, each being an entry point for the user.

1.2.1 Algorithms

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

1.2.2 Tools

Jdemetra+ offers 3 kind of tools

1.2.3 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 buliding blocks of the algorithms presented above, not aiming at all at their comprehensive coverage.

1.3 How to use this book

audience: book targets the beginner as well as seasoned (pun moethodlogist. The beginner is advised to use the quick start chapter as an etrey point, it's presneted like a decsion tree which will point directly to the the part it's useful to review.

the seasoned methodologist will benefit from the detailed chapter ad part strucutre to quickly find the needed information.

2 Quick start with...

objective: describe key steps + provide useful liks to relevant code

2.1 Seasonal Adjustment

2.2 Seasonal Adjustment of High-Frequency Data

2.3 Use of JD+ algorithms in R

2.4 Use of JD+ graphical interface

3 Main functions overview

link to key references * 2 handbooks * sets of guidelines

Objective: present JDemetra+ capabilities by category

3.1 Seasonal adjustment algorithms

below: pieces of old pages to edit and update

3.1.1 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.

3.1.2 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, and the SEATS program developed at the Banco de España. The program is now used by the U.S. Census Bureau for a seasonal adjustment of time series.

3.1.3 Tramo-Seats

3.1.4 STL

3.1.5 Basic Structural Models

3.2 Trend-cycle estimation

3.3 Nowacsting

3.4 Temporal Disaggregation

4 Seasonal Adjustment

4.1 Motivation

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

5 Outlier detection

in or outside a seasonal adjustment process

5.1 Motivation

5.2 With Reg Arima models

5.2.1 Part of preadjustment

5.2.2 Specific TERROR tool

5.3 With structural models

6 Calendar correction and user-defined corrections

generating Calendar regressors and other input variables

6.1 Motivation

6.2 Underlying theory

6.2.1 Overview of Calendar effects in JDemetra+

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.

7 Seasonal adjustment of high frequency data

7.1 Motivation

7.1.1 Ubiquitous use

7.1.2 Data specificities

7.2 Tools

code here and/or link to R packages chapter

7.3 Unobserved Components

7.4 Seasonality tests

7.4.1 Spectral analysis

7.5 Pre-adjustment

7.5.1 Calendar correction

7.5.2 Outliers and intervention variables

7.5.3 Linearization

7.6 Decomposition

7.6.1 Extended X-11

7.6.2 STL decomposition

7.6.3 Arima Model Based (AMB) Decomposition

7.6.4 State Space framework

7.7 Quality assessment

7.7.1 Residual seasonality

7.7.2 Residual Calendar effects

7.7.3 Arima Model

7.8 Forecasting

8 Benchmarking and temporal disagreggation

8.1 Benchmarking overview

Often one has two (or multiple) datasets of different frequency for the same target variable. Sometimes, however, these datasets are not coherent in the sense that they don't match up. Benchmarking[^1] is a method todeal 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.

8.2 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.

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

[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).

$$\left\{x_{i,t}\right\}_{i\in I}$$

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 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 initialiser improves in a significant way the stability of the algorithm.

[5]CHOLETTE, P.A. (1979).

8.3 Tools

8.4 References

9 Trend-cycle estimation

9.1 Motivation

9.2 Underlying Theory

9.3 Tools

9.3.1 rjdfilters package

10 Nowcasting

10.1 Motivation

10.2 Underlying Theory

10.3 Tools

11 Graphical User Interface

11.1 Overview

why use the graphical user interface ? what is not directly available in R yet?

objective: describe the general features (independent of algorithms) - general layout - import data - documents - workspaces - specifications - output

old content can be recycled but - very heavy (trim from md or txt files) - check version 3 - see if we stick with pasted screen shots

12 R packages

12.1 Available algorithms

table

12.2 Organisation overview

a suite (order)

general output organisation

12.3 Installation procedure

12.4 Interaction with GUI

12.5 Full list

12.5.1 rjd3modelling

main functions: table

13 Plug-ins for JDemetra+

13.1 Main functions

table

14 Production

14.0.1 Revision Policies

obj here: general explanations + examples ? here : explain voc discrepancies vs guidelines bbk controlled current link to plug in illustration links on covid

JDemetra+ offers several options for refreshing the output, which are in line with the ESS Guidelines on Seasonal Adjustment (2015) (link) requirements.

reprduce table cf. my pdf (xls) doc remark: rjwsa cruncher vignette is not up to date

15 Tool selection issues

might be integrated to another chapter

objective : select wisely between GUI(+ cruncher and plug ins) and R packages

16 Spectral Analysis Principles and Tools

16.1 Chapter building process

16.2 Spectral analysis concepts and overview

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 $f( )$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 $e^{- i } = cos - isin ,$the spectrum can be also expressed as in equation [3].

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