Seasonal Adjustment of Infra-Monthly Time Series

ICES Short Course

Anna Smyk (Insee), James Livsey (US Census Bureau)

Glasgow, June 17th 2024

INTERNATIONAL CONFERENCE ON ESTABLISHMENT STATISTICS
GLASGOW, SCOTLAND • JUNE 17–20, 2024

Outline

- Characteristics of high-frequency data
- Seasonal adjustment specificities
- Tailoring classic algorithms
- Conclusion and Useful Links

High-frequency data

High-frequency data: usually refers to time series with a frequency higher than monthly

- weekly (ex: traffic casualties)
- daily (daily births, deaths)
- hourly (electricity consumption)

Motivation

High-frequency data:

- becomes ubiquitous in official statistics (digital transformations of data collection give access to infra-monthly economic data and covid-19 pandemic outbreak was a demand accelerator)
- can be seasonal and hence needs to be seasonally adjusted

Goal in this course: show how seasonal adjustment algorithms developed for monthly and quarterly series had to be modified (in JDemetra+ v3) for dealing with HF data

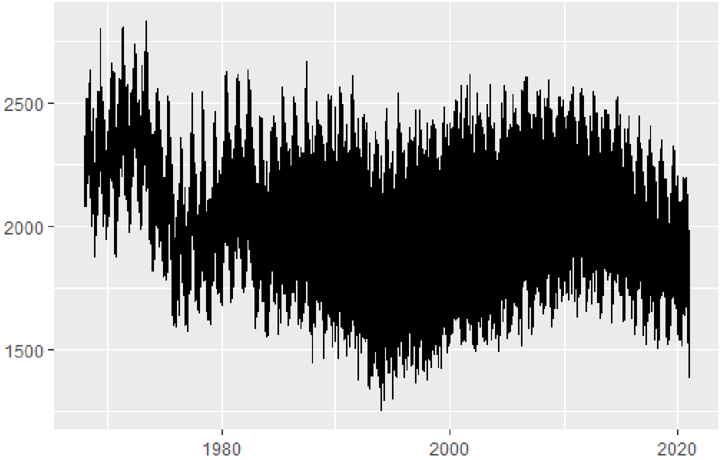
Seasonal adjustment steps (quick reminder)

- seasonality identification (plots, tests)
- linearization (removing outliers and calendar effects)
- decomposition (moving averages, model based...)
- computation of the final series sa series
- testing for residual seasonality

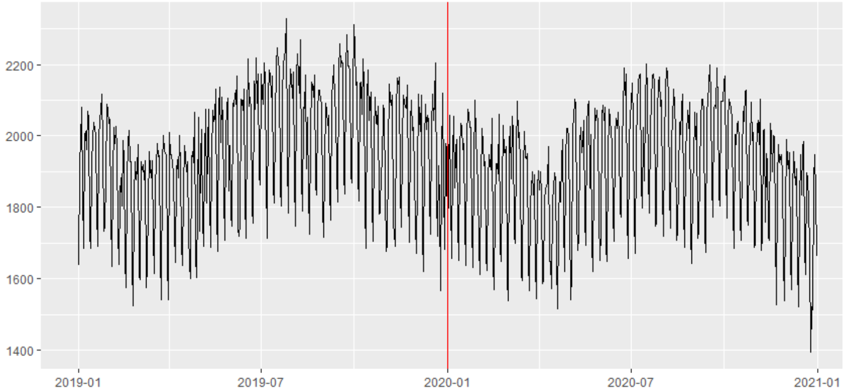
Section 1

Characteristics of high-frequency data

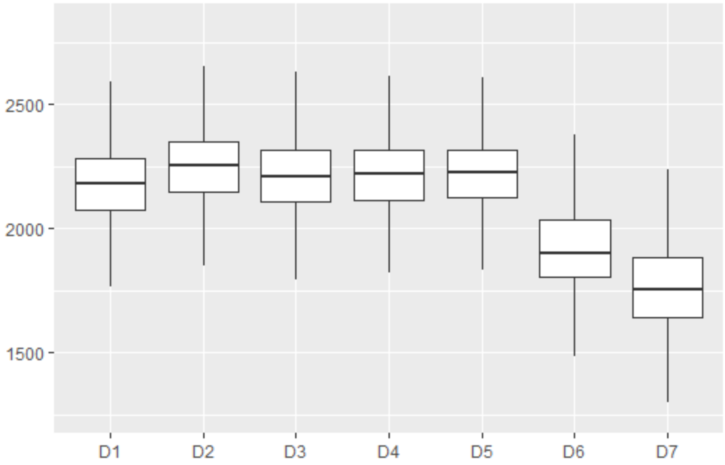
Example: Daily births in France 1968-2020



Example: Daily births in France zoom 2019-2020

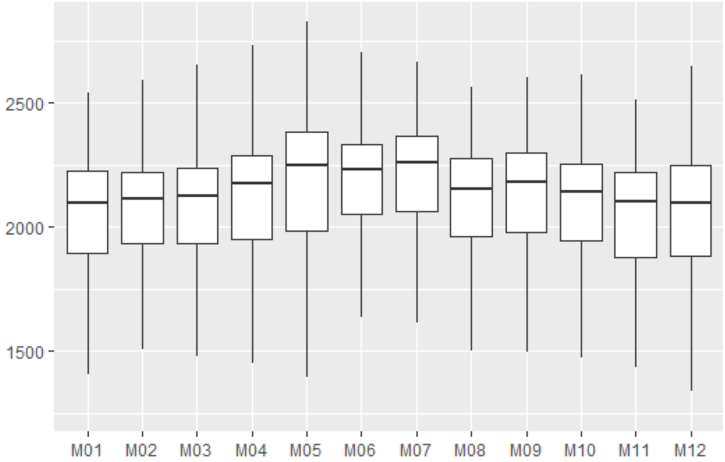


Daily births in France broken down by day of week (1968-2020)



Highlighting weekly periodicity ($p = 7$)

Daily births in France broken down by month (1986-2020)



Highlighting yearly periodicity ($p = 365.25$)

Infra-yearly periodicities : multiple and non integer (1/2)

High-frequency data can display **multiple** and **non integer** periodicities

	periodicities (number of observations par cycle)			
data	day	week	month	year
quarterly				4
monthly				12
weekly			4.348125	52.1775
daily		7	30.436875	365.2425
hourly	24	168	730.485	8765.82

Figure 1: Multiple Periodicities

Infra-yearly periodicities : multiple and non integer (2/2)

A daily series daily might display 3 periodicities

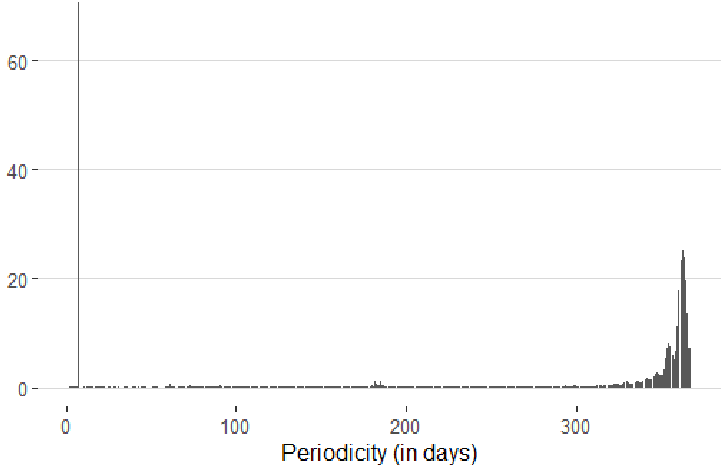
- weekly ($p = 7$): Mondays are alike and different from Sundays (DOW)
- intra-monthly ($p = 30.44$): the last days of each month are different from the fist ones (DOM), much less common than the previous one
- yearly periodicity ($p = 365.25$) : from on year to another the 15th of June are alike, summer days are alike (DOY)

Section 2

Seasonal adjustment specificities

Identification of seasonal patterns

Canova-Hansen test allows to identify multiple seasonal patterns



Decomposition into Unobservable Components

Usual decomposition for seasonal adjustment

$$Y_t = T_t \circ S_t \circ I_t$$

Modification for a daily series: (iterative) estimation of multiple seasonal factors

$$S_t = S_{t,7} \circ S_{t,30.44} \circ S_{t,365.25}$$

If decomposition is Additive ($\circ = +$), if multiplicative ($\circ = \times$)

Seasonality and calendar effects (1/2)

Structural calendar effects

- disturb the comparison between two similar periods
- this is will be modelled as a deterministic effect and corrected by regression in the pre-adjustment phase

The definition of a calendar effect depends on data granularity

For monthly or quarterly data: mostly trading days effect

- regressors are numbers of days of a given type in contrast to Sundays + holidays

Seasonality and calendar effects (2/2)

For daily series

still need to remove holidays effect to make days of a given type comparable

- when estimating S_7
- when estimating $S_{365.25}$

The effect of fixed holidays can be directly allocated to $S_{365.25}$ or corrected as calendar effect

- regressors are vectors of 0's and 1's for a given holiday each year

Section 3

Tailoring classic algorithms

Tailoring classic algorithms

Classic seasonal adjustment algorithms, designed for monthly or quarterly data, cannot tackle multiple and fractional periodicities

Two classes of solutions for fractional periodicities :

- use a Taylor approximation for fractional powers of the backshift operators ($B^{s+\alpha} \approx (1-\alpha)B^s + \alpha B^{s+1}$, see below)
- round periodicities

Decomposition might be done iteratively periodicity by periodicity starting with the smallest one (highest frequency) as:

- highest frequencies usually display the biggest and most stable variations
- cycles of highest frequencies can mix up with lower ones

Seasonal Adjustment Algorithms Key Modifications

Task	Genuine Principle	Key Modification
Pre-Adjustment	Reg-Arima modelling	Fractional powers (backshift operator), multiple seasonal patterns in Airline Model
SEATS	AMB Decomposition	Fractional powers (backshift operator), multiple seasonal patterns in Airline Model
X-11 decomposition	Moving average based sequential trend-cycle and seasonal extraction	Fractional powers (backshift operator), kernel-based trend-cycle filters, iterations on multiple seasonal patterns
STL	Loess filters	Rounding down fractional periodicities, iterations on multiple seasonal patterns
STS	Explicit modelling of components, one-step	Rounding down fractional periodicities

In this course focus on Reg-Arima modelling and X-11 decomposition

Series linearization with Reg-Arima model

In X13-Arima and Tramo-Seats

- Reg-Arima modelling step
- to remove deterministic effects: outliers and calendar
- outliers will be re-injected into the SA series

The Reg-ARIMA model is written as follows:

$$\left(Y_t - \sum \alpha_i X_{it}\right) \sim ARIMA(p,d,q)(P,D,Q)$$

These models contain seasonal backshift operators $B^s(y_t) = y_{t-s}$

Modification of the Airline model (1/2)

"Airline" model is $ARIMA(0,1,1)(0,1,1)$:

$$(1-B)(1-B^s)y_t = (1-\theta_1 B)(1-\theta_2 B^s)\epsilon_t \quad \epsilon_t \sim \text{NID}(0,\sigma_\epsilon^2)$$

for high-frequency data:

- the model might contain several differentiations $\Delta_s = 1 - B^s$ and also B^s with non integer s
- we write $s = s' + \alpha$, with α real number i $]0,1[$ (for example $52.18 = 52 + 0.18$ is the yearly periodicity for weekly data)

Modification of the Airline model (2/2)

With a Taylor development around 1 of $f(x) = x^\alpha$

$$\begin{array}{rcl} x^\alpha & = & 1 + \alpha(x-1) + \frac{\alpha(\alpha+1)}{2!}(x-1)^2 + \frac{\alpha(\alpha+1)(\alpha+2)}{3!}(x-1)^3 + \cdots \\ B^\alpha & \cong & (1-\alpha) + \alpha B \end{array}$$

Approximation of $B^{s+\alpha}$: fractional Airline model

$$B^{s+\alpha} \quad \cong \quad (1-\alpha)B^s + \alpha B^{s+1}$$

Modelling of daily births series

Two periodicities $p_1 = 7$ and $p_2 = 365.25$

$$(1-B)(1-B^7)(1-B^{365.25)}(Y_t - \sum \alpha_i X_{it}) = (1-\theta_1 B)(1-\theta_2 B^7)(1-\theta_3 B^{365.25})\epsilon_t$$

$$\epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2)$$

with

$$1 - B^{365.25} = (1 - 0.75B^{365} - 0.25B^{366})$$

Extension of X-11 decomposition (1/2)

Moving average based decomposition

- global structure of iterations identical to genuine X-11
- iterative decomposition done periodicity by periodicity starting with the smallest one
- extension of the preliminary trend filter definition for removing seasonality
- extension of final trend estimation filters
  - genuine X-11: Henderson filters (+ Musgrave asymmetrical surrogates)
  - extended X-11: generalization of this method with local polynomial approximation (different weight distributions)
- modification of the seasonality extraction filters to take into account fractional periodicities

Modification of the first trend filter for removing seasonality

For the first trend estimation: generalization of centred and symmetrical moving averages with an order equal to the periodicity p;

- filter length longueur l: smallest odd integer greater than p
- ex : p=7, l=7, p=12 l=13, p=365,25, l=367, p=52.18 l=53
- central coefficients $1/p$ (1/12,1/7, 1/365.25)
- extreme coefficients $\mathbb{I}\{E(p)$ pair$\} + (p - E(p))/2p$
- ex : p=12 (1/12 and 1/24) (we fall back on $M_{2\times 12}$ of the monthly case)
- ex : p=365.25 (1/365.25 and 0.25/(2*365.25)

Modification of seasonality extraction filters (1/2)

Computation is done on a given period

Example $M_{3\times 3}$

$$M_{3\times 3}X = \frac{1}{9}(X_{t-2p}) + \frac{2}{9}(X_{t-p}) + \frac{3}{9}(X_t) + \frac{2}{9}(X_{t+p}) + \frac{1}{9}(X_{t+2p})$$

if p integer, nothing to change if p non integer we use the Taylor approximation of the backshift operator

$$B^{s+\alpha} \quad \cong \quad (1-\alpha)B^s + \alpha B^{s+1}$$

Modification of seasonality extraction filters (2/2)

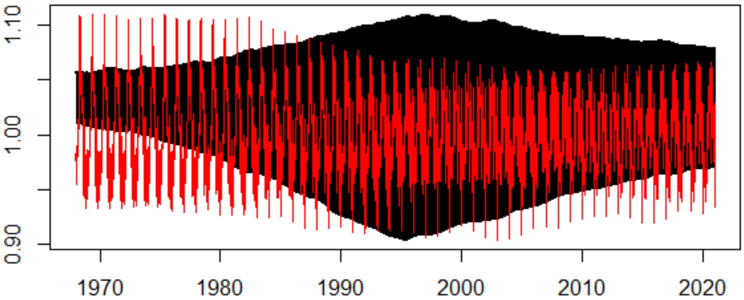
For example, for $p = 30.44$ filter 3×3 is written as follows:

$$
\begin{aligned}
\hat{s}_t &= \frac{1}{9}\left[0.88 \times \widehat{(si)}_{t-61} + 0.12 \times \widehat{(si)}_{t-60}\right] \\
&+ \frac{2}{9}\left[0.44 \times \widehat{(si)}_{t-31} + 0.56 \times \widehat{(si)}_{t-30}\right] \\
&+ \frac{3}{9}\widehat{(si)}_{t} \\
&+ \frac{2}{9}\left[0.56 \times \widehat{(si)}_{t+30} + 0.44 \times \widehat{(si)}_{t+31}\right] \\
&+ \frac{1}{9}\left[0.12 \times \widehat{(si)}_{t+60} + 0.88 \times \widehat{(si)}_{t+61}\right]
\end{aligned}
$$

This approximation avoids data imputation

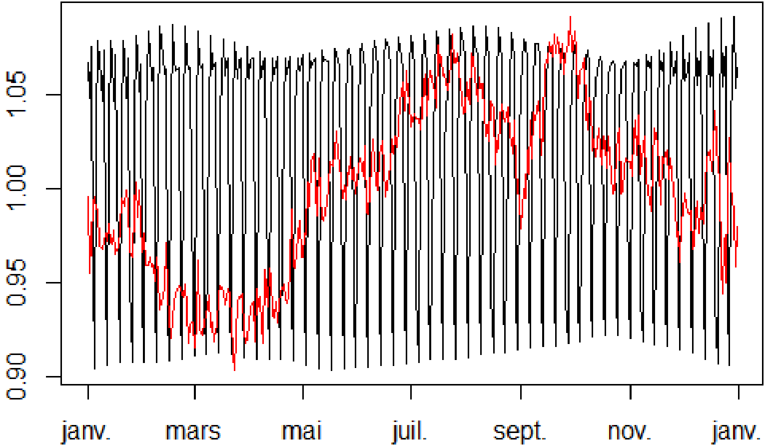
Daily births in France (1968-2020): final seasonal patterns (1/2)

Evolving seasonal patterns computed on a linearized series (outliers and calendar effects removed)

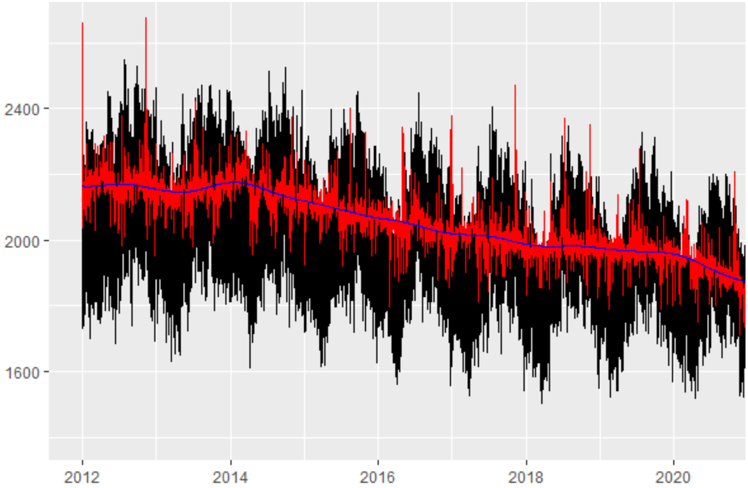


Daily births in France (1968-2020): final seasonal patterns (2/2)

Zoom on the year 2000: estimated seasonal factors: $p = 7$ (black) and $p = 365.25$ (red)

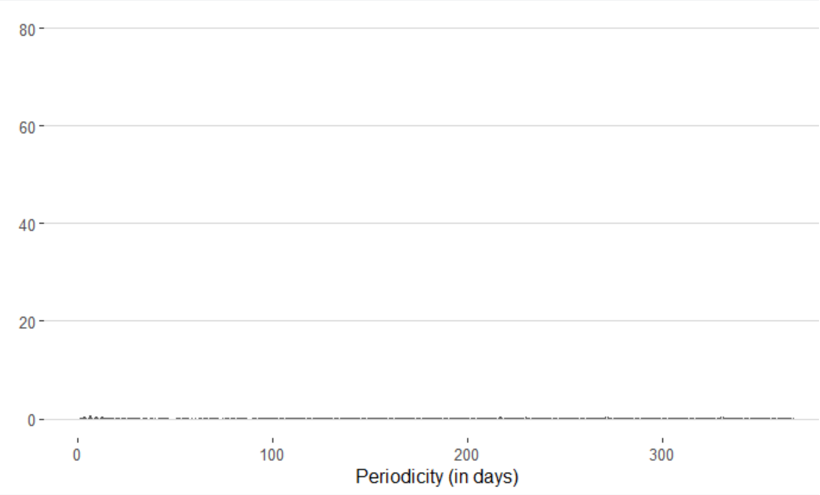


Daily births : raw, sa and trend



Residual seasonality

Canova-Hansen test on final SA series estimated with extended X-11



Daily births in France (1968-2020): final results in JDemetra+ GUI

Results of Extended airline linearization and Decomposition (SEATS) in the Graphical User Interface

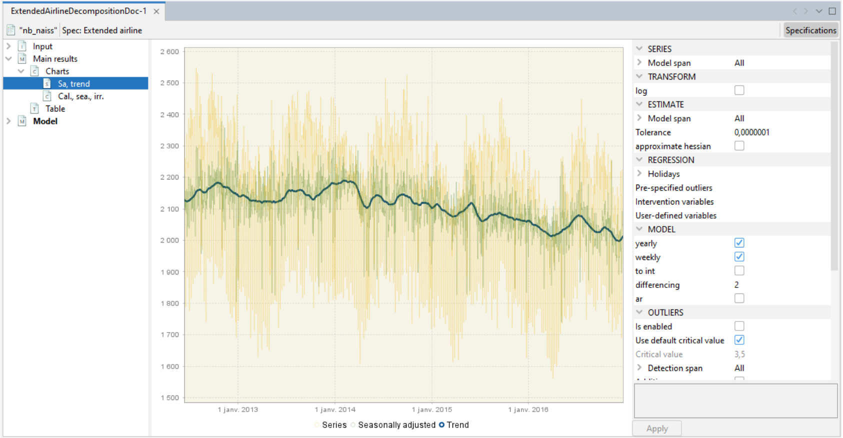


Figure 4: Raw, SA and trend

Algorithms in R

Available R packages (based on JDemetra+) for SA of high frequency data

Pre-treatment

- Extended Airline Model in `rjd3highfreq`

Decomposition

- Extended SEATS in `rjd3highfreq`
- Extended X-11 in `rjd3x11plus`
- Extended STL `rjd3stl`

One-Step SA with explicit components

- SSF Framework `rjd3sts`

Section 4

Conclusion and Useful Links

Conclusion

Main challenges when seasonally adjusting high-frequency data: multiple and non integer periodicities

Classic algorithms have been (partly) tailored to this purpose

On going investigations (around JDemetra+ v3.x algorithms)

- Seasonal factor estimation: cubic splines for p=365.25
- Automatic filter selection (X-11, STL)
  - Trend-cycle filters: modified I/C ratio? cross validation ? Kernel Parameters ?
  - Seasonal filters: Modified I/S ratio? Window length? Spectral approaches?
- Model extensions (pre-treatment and AMB)
  - Arima orders: beyond airline?
  - Fractional periodicities: beyond Taylor approximation ?

Useful Links

- Github repository for the course https://github.com/jlivsey/ICES2024-timeSeries
- Towards Seasonal Adjustment of Infra-Monthly Time Series with JDemetra+, Webel and Smyk (2023), Bundesbank Discussion Paper
- R Packages giving access to JDemetra+ v3.x: https://github.com/rjdverse
- Graphical User Interface version 3.x: https://github.com/jdemetra/jdplus-main
- JDemetra+ Online documentation: https://jdemetra-new-documentation.netlify.app/

Software Demonstration on JDemetra+ at this conference on Wednesday 19th, 12:45.