Towards Seasonal Adjustment of Infra-Monthly Time Series with JDemetra+ 3.0

rjd3highfreq R package

Anna Smyk[1] Karsten Webel[2]

[1]Institut National de la Statistique et des Études Économiques (INSEE)

[2]Deutsche Bundesbank (BBk)

uRos Virtual Conference, Hosted by Statistics Canada, December 6th 2022

Table of Contents

High-frequency data

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

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

These series can be seasonal and become more and more ubiquitous in official statistics

Goal of this presentation

- show how seasonal adjustment algorithms developed for monthly and quarterly series had to be modified in JDemetra+ v3.0 (SA software) for dealing with HF data

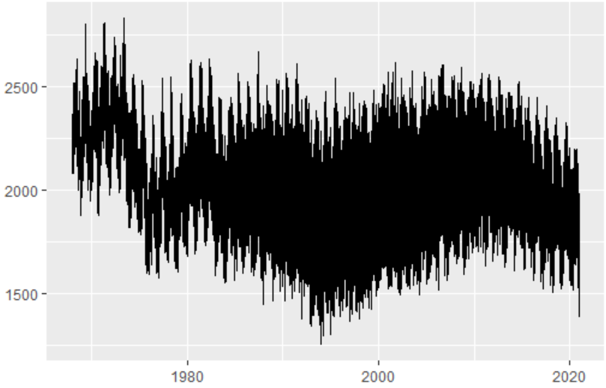
These algorithms are available within the {rjd3highfreq} R package

Goal of seasonal adjustment

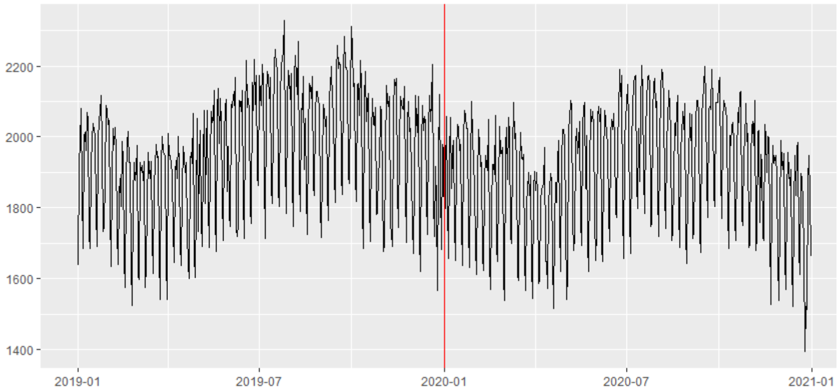
- remove from the series infra-yearly periodic movements
- requires estimating seasonal (S) and calendar (C) factors to be subtracted from the raw series $Y_{sa} = Y - S - C$
- for this estimation, the series will be split into unobservable components: seasonal, trend and irregular ($Y = T + S + I$), after having been corrected for C
- two algorithms are widely used in official statistics: X13-Arima and Tramo-Seats

Plan

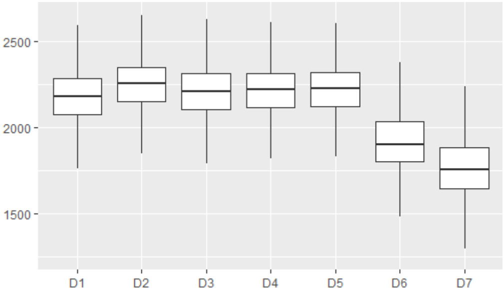
Daily births in France 1968-2020



Daily births in France zoom 2019-2020

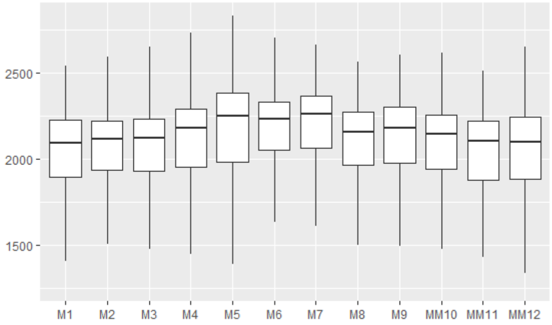


Daily births in France broken down by day of week (1968-2020)



Highlighting weekly periodicity ($p = 7$)

Daily births in France broken down by month (1968-2020)



Highlighting yearly periodicity ($p = 365.25$)

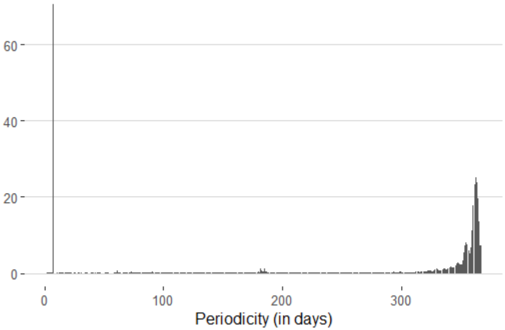
Multiple and non integer periodicities

	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

a daily series 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 first ones (DOM)
- yearly ($p = 365.25$) : from on year to another the 15th of June are alike, summer days are alike (DOY)

Canova-Hansen test



Decomposition into unobservable components

series raw	$=$	trend	$\circ$	seasonal	$\circ$	calendar	$\circ$	irregular
$\downarrow$		$\downarrow$		$\downarrow$		$\downarrow$		$\downarrow$
y_t	$=$	t_t	$\circ$	s_t	$\circ$	c_t	$\circ$	i_t
		$\downarrow$		$\downarrow$		$\downarrow$		$\downarrow$
		Long-term variations		Infra-yearly periodic variations		Calendar		Residual
		$\searrow$						$\swarrow$

seasonally adjusted (sa) data

- Decomposition: Additive ($\circ = +$), multiplicative ($\circ = \times$)

Multiple seasonal factors

New equation for high-frequency data:

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

decomposition will 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

Seasonality and calendar effects

- calendar effects disturb the comparison between two periods, their definition depends on the data frequency and on the periodicity under review
- for daily series:
  - remove the bank 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 in the pre-adjustment phase

Plan

Linearization

In X13-Arima and Tramo-Seats

- Reg-Arima modeling 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 $B^s(y_t) = y_{t-s}$

The fractional Airline model

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

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

for high frequency-data:

- the model might contain several differentiation $\Delta_s = 1 - B^s$ and thus B^s with non integer s
- we write $s = s' + \alpha$, with α real number in $]0,1[$ (for example $52.18 = 52 + 0.18$ for weekly data)
- and use a Taylor approximation of $B^{s+\alpha}$

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

Modeling of daily births series

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

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

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

with

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

Linearization: parameter selection

```
pre.mult<- rjd3highfreq::fractionalAirlineEstimation
                        (df_daily$log_births, # here series in log
                x = q, # q= calendar
                periods = 7, # approx  c(7,365.25)
                ndiff = 2, ar = FALSE, mean = FALSE,
                outliers = c("ao","wo"),
                # WO compensation, LS not relevant here
                criticalValue = 0, # computed in the algorithm
                precision = 1e-9, approximateHessian = TRUE)

# calendar regressors can be defined with the {rjd3modelling} package
```

Linearization results : calendar effects

Variable	Coef	Coef_SE	Tstat
14jt	-0.12	0.00	-26.00
8mai	-0.15	0.01	-28.71
asc	-0.17	0.00	-38.72
01jan	-0.26	0.00	-52.87
e_mon	-0.19	0.00	-42.44
1mai	-0.12	0.00	-24.81
l_pen	-0.19	0.00	-42.64
15aou	-0.12	0.00	-26.29
1nov	-0.15	0.00	-33.75
11nov	-0.13	0.00	-27.52
25dec	-0.28	0.00	-55.91

Linearization results: outliers

Variable	Coef	Coef_SE	Tstat
AO.1993-12-24	-0.19	0.03	-5.77
WO.2001-03-19	0.12	0.02	5.72
AO.1995-08-14	-0.19	0.03	-5.74
AO.1997-08-15	-0.18	0.03	-5.54
AO.1970-12-25	0.19	0.03	5.85
AO.2011-05-08	0.19	0.03	5.75
AO.2018-11-11	0.18	0.03	5.50
AO.2017-01-01	0.23	0.03	7.06
AO.1978-01-01	0.23	0.03	6.93
AO.1997-12-24	-0.18	0.03	-5.49
WO.2006-01-01	0.14	0.02	6.38
AO.1998-05-01	-0.19	0.03	-5.74

Plan

X-11 overview

- X-11 is the decomposition module of X-13-Arima
- the linearized series from the pre-adjustment step is split into S, T, I
- decomposition is done periodicity by periodicity starting with the smallest one
- global structure of the iterations is the same as in genuine X-11
- modifications for tackling non integer periodicities

Modification of the preliminary trend filter for removing seasonality

for the first trend estimation: generalization of centered and symmetrical moving averages with an order equal to the periodicity p

- filter length 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)
- end-point coefficients $\mathbb{I}\{E(p) \text{ even}\} + (p - E(p))/2p$
- 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)
- sum of weights equals one

Modification of the seasonality extraction filters (1/2)

computation is done on a given period
for 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

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

Modification of seasonality extraction filters (2/2)

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

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

this approximation allows avoiding data imputation

Modification of the final trend estimation filter

As seasonality has been removed in the first step, there is no non integer periodicity issue in the final trend estimation, but {rjd3highfreq} offers an extension vs genuine X-11

- genuine X-11: Henderson filters (+ Musgrave asymmetrical surrogates)
- in {rjd3highfreq} generalization of this method: trend-cycle estimation according to Proietti and Luati (2008) founded on applying local polynomial regressions to the input series

Extended X-11 for p=7: parameters

```
x11.dow <- rjd3highfreq::x11(exp(pre.mult$model$linearized),
        period = 7,                 # DOW pattern
        mul = TRUE,
        trend.horizon = 9,  # 1/2 Filter length : not too long vs p
        trend.degree = 3,                         # Polynomial degree
        trend.kernel = "Henderson",               # Kernel function
        trend.asymmetric = "CutAndNormalize",     # Truncation method
        seas.s0 = "S3X9", seas.s1 = "S3X9",       # Seasonal filters
        extreme.lsig = 1.5, extreme.usig = 2.5)   # Sigma-limits
```

Extended X-11 for p=365.25 : parameters

```
x11.doy <- rjd3highfreq::x11(x11.dow$decomposition$sa,  # previous sa
                    period = 365.2425,         # DOY pattern
                    mul = TRUE,
                    trend.horizon = 371, # 1/2 final filter length
                    trend.degree = 3,
                    trend.kernel = "Henderson",
                    trend.asymmetric = "CutAndNormalize",
                    seas.s0 = "S3X15", seas.s1 = "S3X5",
                    extreme.lsig = 1.5, extreme.usig = 2.5)
```

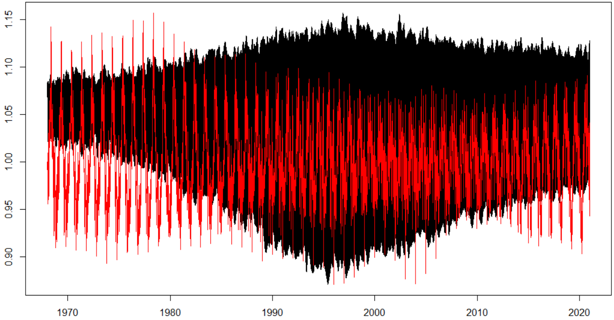
Missing: criteria for filter selection and correction thresholds

Summary of the SA process

- linearization : $Y_{lin} = FracAirline(Y)$, computation of Y_{cal}
- $sa_7 = X11_7(Y_{lin})$, computation of S_7
- $sa_{365.25} = X11_{365.25}(sa_7)$, computation of $S_{365.25}$
- $sa_{final} = Y_{cal}/S_7/S_{365.25}$

Decomposition of daily births series (1/2)

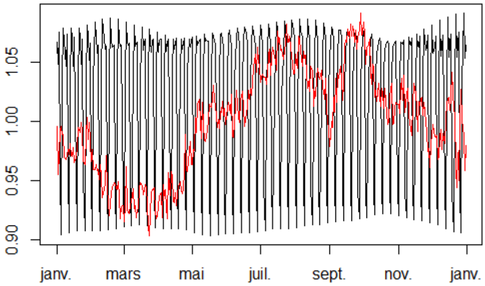
estimated seasonal factors : p=7 (black) and p=365.25 (red)



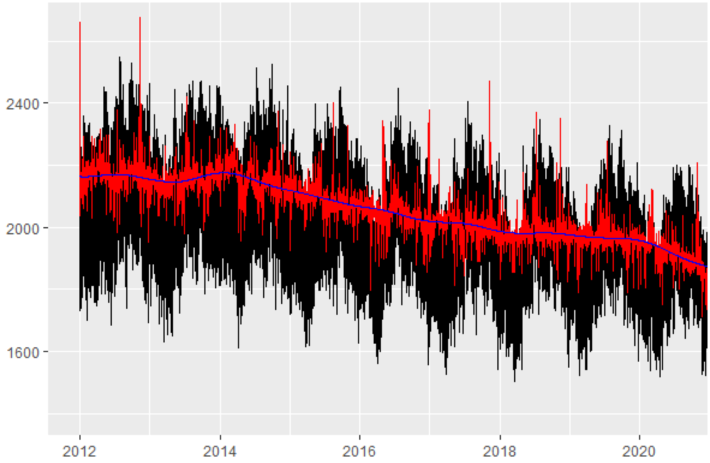
evolving seasonality over a long period

Decomposition of daily births series (2/2)

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



Daily births : raw, sa and trend



Plan

Extended SEATS

Arima model based decomposition (AMB, extension of Seats) also available in {rjd3highfreq}
Parameters:

```
amb.doy <- rjd3highfreq::fractionalAirlineDecomposition(
  amb.dow$decomposition$sa,  # DOW-adjusted linearised data
  period = 365.2425,         # DOY pattern
  sn = FALSE,                # Signal (SA)-noise decomposition
  stde = FALSE,              # Calculate standard deviations
  nbcasts = 0, nfcasts = 0)  # Numbers of back- and forecasts
```

AMB filters are optimal by design

Results slightly differ from X-11 (just as for monthly or quarterly data)

Plan

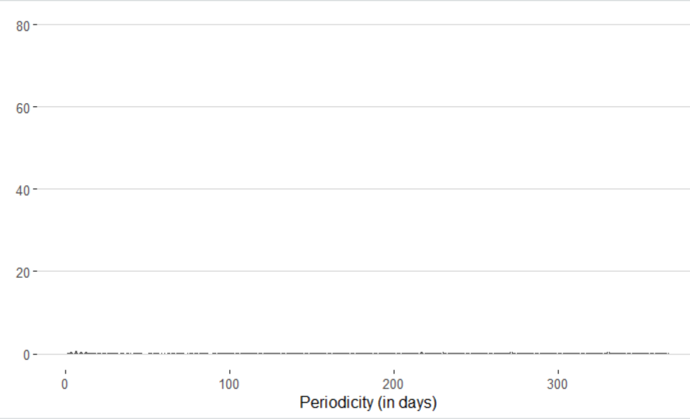
Residual seasonality (1/2)

We should check that the final SA series doesn't display any residual seasonality, for all periodicities under review

- usual seasonality tests cannot always be used in the high-frequency data framework (ex : if Anova based, multiple and non integer periodicities are a problem)
- we use again Canova-Hansen test based on the spectrum

Residual seasonality (2/2)

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



same scale as the test on raw series

Plan

Conclusion

- main challenges when seasonally adjusting HF data: multiple and non integer periodicities
- JDemetra+ v3.0 features several tailored algorithms: X-13-Arima, Tramo-Seats, mentioned here, but also STL, all available in the {rjd3highfreq} R package
- further research: developing seasonality tests and honing criteria for filter and parameter selection

Thank you for your attention

- {rjd3highfreq} : `https://github.com/palatej`