"X11-Like" Seasonal Adjustment of Daily Data

D. Ladiray, J. Palate, G.L. Mazzi, T. Proietti [1]

CARMA 2018, Valencia

[1]Contact email: dominique.ladiray@insee.fr

Outline

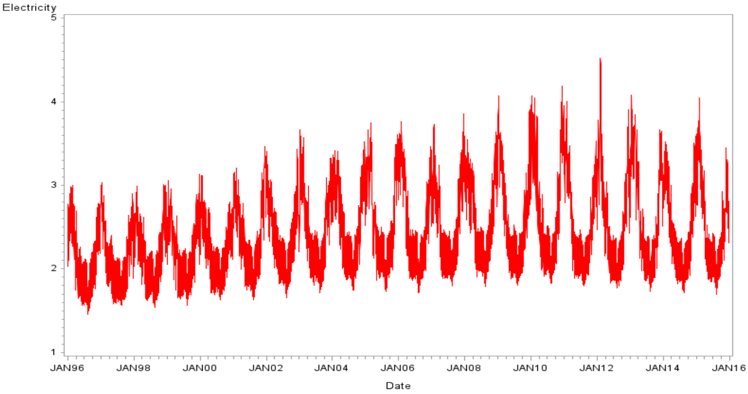
Introduction

- A joint work:
  - Dominique Ladiray (INSEE), Gian Luigi Mazzi (GOPA), Jean Palate (NBB), Tommaso Proietti (University of Rome Tor Vergata)
  - Complete paper published in the "Handbook on Seasonal Adjustment" (Eurostat, May 2018);
    - `ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001`
- On the agenda of the European Seasonal Adjustment Center: improving SA methods in JDemetra+, the European time series software for official statistics.
  - Extending "X11", STL and Tramo-Seats to daily and weekly data;
  - Incorporating State Space Methods;
  - New filters for "X11" (Dagum's asymmetric filters)

Time Series Software
for Official Statistics

Example

Daily consumption of electricity in France since 01/01/1996.
Clear annual pattern.



Multiples periodicities

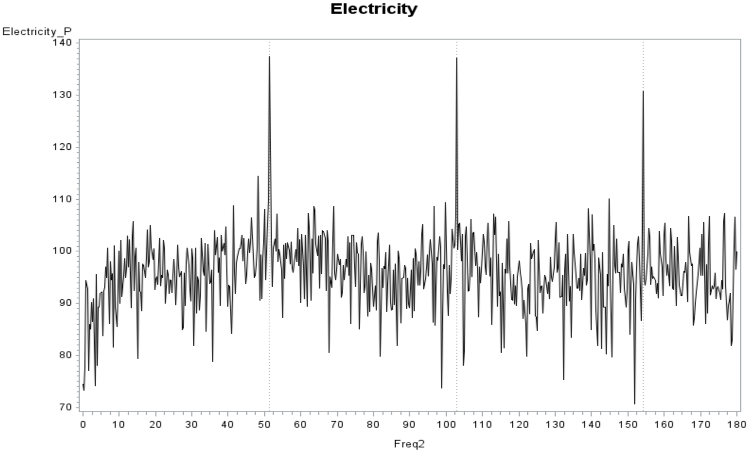
Data	Minute	Hour	Day	Week	Month	Quarter	Year
	Period (number of observations per cycle)						
Annual							1
Quarterly							4
Monthly						3	12
Weekly					4.348125	13.044375	52.1775
Daily				7	30.436875	91.310625	365.2425
Hourly			24	168	730.485	2191.455	8765.82
Half-hourly			48	336	1460.97	4382.91	17531.64
Minutes		60	1440	10080	43829.1	131487.3	525949.2
Seconds	60	3600	86400	604800	2629746	7889238	31556952

A daily series might have 3 different (and hopefully co-prime) periodicities:

- A weekly periodicity (7 days): Monday is different from Sunday;
- An intra-monthly periodicity (30.436875 days on average): the first days of the month are different from the last days;
- An annual periodicity (365.2425 days on average): each day if different from others (Winter days are different from Summer days).

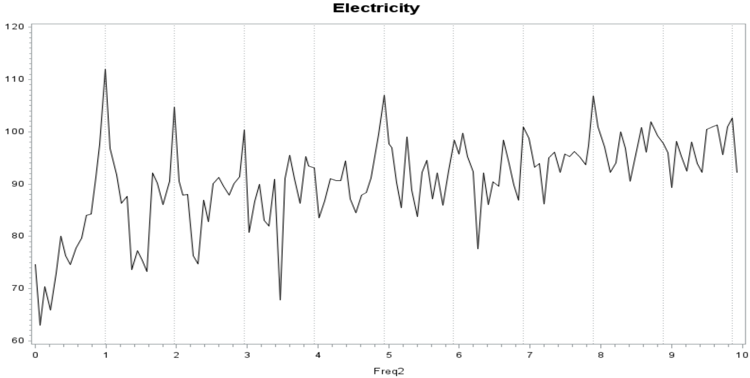
Periodogram of the daily consumption of electricity.

Peaks at harmonics $2k\pi/7$. Annual periodicity is hidden.



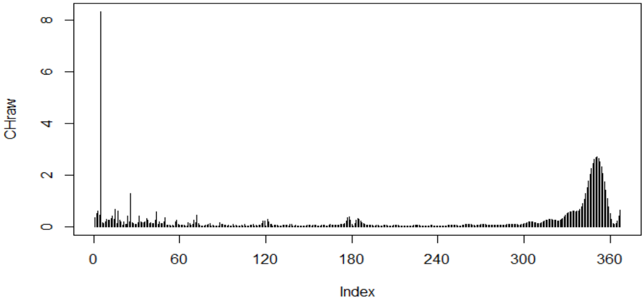
Zoom on the low frequencies

Peaks at harmonics $2k\pi/365$.



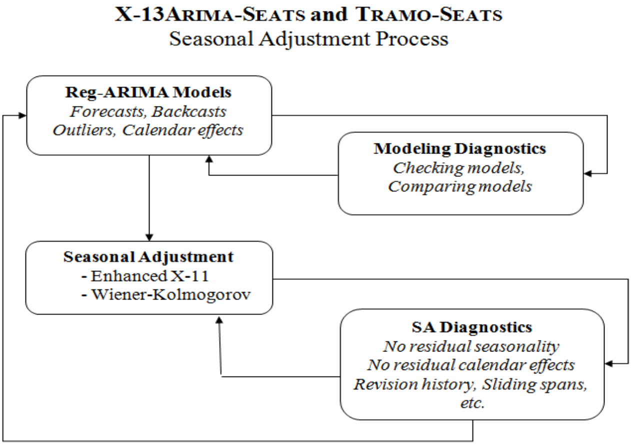
Canova-Hansen Test

Canova-Hansen tests for components with period $p = 2, 3, \ldots, 366$.



Significant values are detected at the weekly and annual frequencies only, for which the null is rejected.

Usual Methodology



The ARIMA model (1)

- ARIMA models $(p, d, q)(P, D, Q)_S$ with one seasonality and first or/and seasonal integer differences $\Delta = 1 - B$ or/and $\Delta_s = 1 - B^s$.
- But we now have several and possibly non integer periodicities. $s + \alpha$ is the periodicity (for example 365.25 for daily data or 52.18 for weekly data) where s is the integer part and α a real number belonging to the interval $]0, 1[$.
- How can we adapt ARIMA models to this particular situation?

The ARIMA model (2)

- Using the Taylor expansion of x^α, we have:

$$\begin{aligned} x^\alpha &= 1 + \alpha(x-1) + \tfrac{\alpha(\alpha+1)}{2!}(x-1)^2 + \tfrac{\alpha(\alpha+1)(\alpha+2)}{3!}(x-1)^3 + \cdots \\ &\cong (1-\alpha) + \alpha x \end{aligned}$$

- We define the "tilde difference operator":

$$\begin{aligned} \tilde{\Delta}_{s+\alpha} y_t &= y_t - B^{s+\alpha} y_t \\ &= y_t - B^s B^\alpha y_t \\ &\cong y_t - (1-\alpha) B^s y_t + \alpha B^{s+1} y_t \end{aligned}$$

Extensions of the airline model

1. For weekly series:
   $\Delta\Delta_{52} y_t = (1 - \theta_1 B)(1 - \theta_2 B^{52})\epsilon_t$
   $\Delta\tilde{\Delta}_{52.18} y_t = (1 - \theta_1 B)(1 - 0.82\theta_2 B^{52} - 0.18\theta_2 B^{53})\epsilon_t$
2. For daily series:
   $\Delta\Delta_7\Delta_{365} y_t = (1 - \theta_1 B)(1 - \theta_2 B^7)(1 - \theta_3 B^{365})\epsilon_t$
   $\Delta\Delta_7\tilde{\Delta}_{365.25} y_t = (1 - \theta_1 B)(1 - \theta_2 B^7)(1 - 0.75\theta_3 B^{365} - 0.25\theta_2 B^{366})\epsilon_t$
3. And for other periodicities, similar extensions could be considered.

Allows to develop a new "Tramo-like" algorithm for pretreatment (automatic detection of calendar effects and outliers) and forecasting of high-frequency data.

Calendar effects (1)

- In daily data, most of the calendar effects are part of the annual periodicity.
- A few exceptions: moving holidays (Easter, Ramadan etc.) and some fixed holidays (type "first Monday of July").
- Note that looking at daily data can improve the specification of calendar effect models in Monthly data.

Calendar effects (2)

A	B	C	D	E	F	G	H	I	J	K	L
Lag	Jan1	May1	May8	Jul14	Aug15	Nov1	Nov11	Dec25	EasterMor	Ascension	Pentecost
-10	-0,56	-0,463	0	0,529	-1,24	0,449	0,772	0,996	-1,466	1,591	0
-9	-4,706	-1,044	0	0,133	-1,312	0,291	6,508	0,522	-0,747	0,557	0
-8	-15,12	-0,911	0	0,031	-1,243	0,681	-1,878	0,313	0	0,735	0
-7	5,233	-1,469	0,357	0,06	-1,308	0,296	-0,902	-0,311	-0,051	1,011	0
-6	7,448	-1,361	4,193	0,146	-0,855	0,124	1,041	-0,137	-0,164	0,8	0
-5	2,061	-1,115	0,09	0,547	-0,998	0,107	-0,259	-0,493	1,805	1,734	0
-4	-1,055	-0,006	0,057	1,071	-1,263	0,213	-2,543	-0,453	1,516	0	0
-3	-0,154	0,494	0,271	0,909	-1,741	0,93	-1,011	0	-0,054	1,121	0
-2	0,254	1,565	2,229	1,258	-2,591	0,807	3,639	0	1,006	0,971	0
-1	-0,72	-0,601	0,658	-1,123	-3,489	-2,085	1,009	0	0	0,987	0
0	-5,607	-8,336	-7,525	-18,915	-15,032	-8,367	-8,772	-16,212	-23,738	-17,739	0
1	-6,994	-5,271	0	-1,85	-5,225	-6,728	-2,365	-12,495	-2,842	-9,403	-0,36
2	1,217	0,967	0	1,919	-2,613	3,184	0,327	-5,588	-0,938	-1,559	1,648
3	2,738	0,251	0	1,029	-2,5	0,497	1,206	-0,071	0,127	2,126	1,87
4	2,182	-0,064	0,166	1,047	-2,481	-0,294	1,581	-1,774	1,066	0	2,59
5	1,212	-2,424	-0,192	0,933	-2,722	1,314	1,378	-2,363	0,523	2,098	1,349
6	1,085	-1,255	0,177	0,582	-2,348	3,322	1,072	-4,417	0,962	1,97	1,148
7	1,06	1,15	0,844	0,896	-1,788	0,419	0,6	-2,297	0	2,234	0
8	0,796	-3,402	1,13	0,474	-0,842	-4,486	0,703	0	0,098	2,114	1,201
9	0,279	-0,102	0,027	0,593	-0,636	-1,213	0,912	0	-0,573	1,448	0,838
10	1,739	0,002	-0,489	0,355	-0,616	1,57	0,544	0	-0,771	-20,182	0,666

A multi-period decomposition algorithm

1. **Estimation of the Trend-Cycle with a** 7×365
   $C_t^{(1)} = M_{7\times 365}(X_t)$
2. **Estimation of the global seasonal-irregular component**
   $(S_t + I_t)^{(1)} = X_t - TC_t^{(1)}$

3a. **Estimation of the weekly** (7) **seasonal-irregular component (MA** 365)
   $SI_t^{(1),7} = M_{365}\left[(S_t + I_t)^{(1)}\right]$

3b. **Estimation of the weakly seasonal component with a** 3×3
   on each period (day): $S_t^{(1),7} = M_{3\times 3}\left[SI_t^{(1),7}\right]$
   Normalisation : $\tilde{S}_t^{(1),7} = S_t^{(1),7} - M_7\left(S_t^{(1),7}\right)$

3c. **Idem for the annual seasonal component** "365"

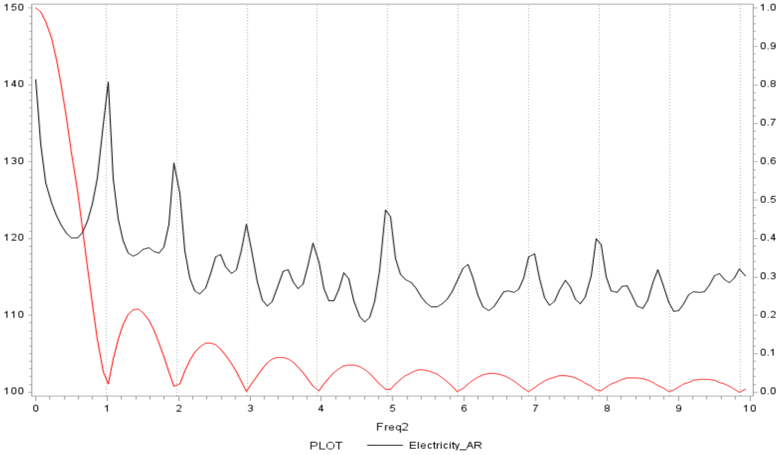
4. **Estimation of the seasonally adjusted series:**
   $A_t^{(1)} = (C_t + I_t)^{(1)} = X_t - \tilde{S}_t^{(1),7} - \tilde{S}_t^{(1),365}$

Seasonal adjustment: Some Issues

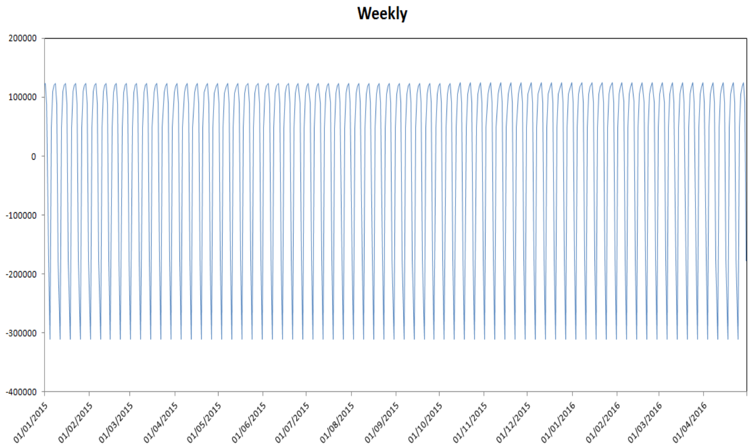
Date	D01	D02	D03	D04		D26	D27	D28	D29	D30	D31
12/2011	0.28	0.40	0.34	0.38	....	0.42	0.34	0.29	0.32	0.30	0.35
01/2012	0.29	0.20	0.09	0.02	....	-0.51	-0.45	-0.48	-0.45	-0.41	-1.21
02/2012	-1.17	-0.29	-0.28	-0.26	....	-0.58	-0.58	-0.57	-0.57	.	.
03/2012	-0.47	-0.40	-0.44	-0.37	....	-0.78	-0.77	-0.73	-0.62	-0.41	-0.45
04/2012	-0.47	-0.41	-0.41	-0.24	....	-0.50	-0.30	-0.30	-0.27	-0.22	.
05/2012	-0.23	-0.21	-0.15	-0.16	....	-0.04	-0.10	-0.07	-0.14	-0.05	0.03
06/2012	0.11	0.10	0.06	0.02	....	-0.26	-0.22	-0.05	0.11	0.15	.
07/2012	0.08	0.08	0.05	0.04	....	0.18	0.27	0.26	0.20	0.26	0.28
08/2012	0.35	0.45	0.48	0.48	....	0.27	0.15	0.13	0.17	0.39	0.49
09/2012	0.47	0.47	0.48	0.38	....	0.32	0.48	0.61	0.63	0.62	.
10/2012	0.73	0.77	0.77	0.78	....	0.34	0.35	0.37	0.39	0.43	0.49
11/2012	0.51	0.59	0.61	0.56	....	-0.16	-0.16	-0.09	0.11	0.23	.
12/2012	0.23	0.24	0.20	0.20	....	0.97	1.01	1.12	1.12	1.09	1.06
01/2013	1.09	0.94	0.77	0.68	....	-0.49	-0.51	-0.53	-0.52	-0.45	-0.27
02/2013	-0.22	-0.24	-0.25	-0.29	....	-0.74	-0.68	-0.54	.	.	.
03/2013	-0.43	-0.49	-0.50	-0.54	....	-0.64	-0.40	-0.07	-0.07	-0.09	-0.09

- Intra-monthly seasonality: the "ragged matrix" problem (28, 29, 30 or 31 days); using some kind of "time warping";
- Leap Year effect: not a big deal (ignore and then estimate by splines).

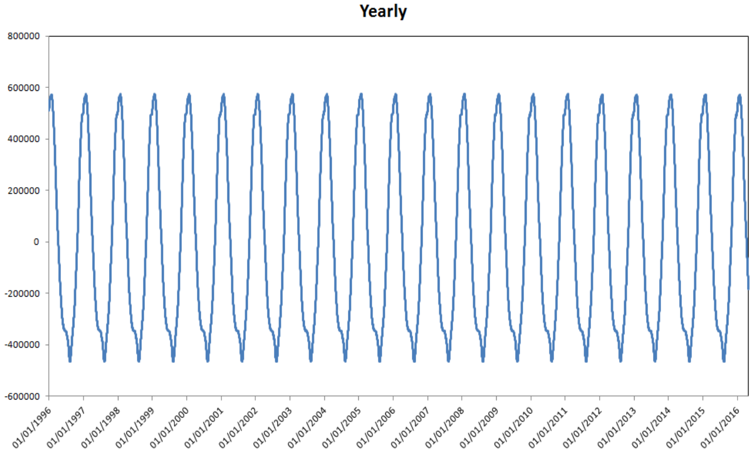
Spectrum and Gain Function



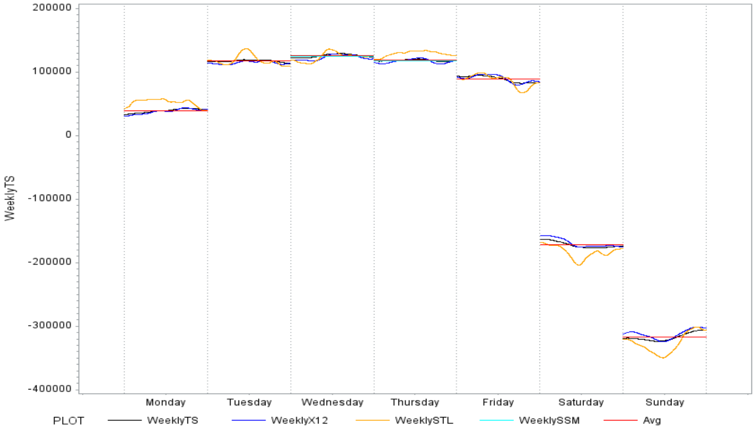
Weekly seasonal component



Yearly seasonal component



X12, SEATS, STL and SSM Weekly Components



First Conclusions

- It is (certainly) possible to adapt the X11 algorithm to daily (and weekly) time series;
- To separate the annual periodicity from the trend-cycle low frequencies might be tricky;
- The X11 iterative philosophy is very close to STL one; and STL can also be adapted to high frequency data;
- Calendar effects (moving holidays for daily series) must be modeled very carefully;
- Still lots of things to tune from this first prototype:
  - moving-average orders;
  - automatic choice of moving-averages (IoverC and IoverS ratios).