Outliers and Calendar Effects

ESTP Training
11-13 October 2022

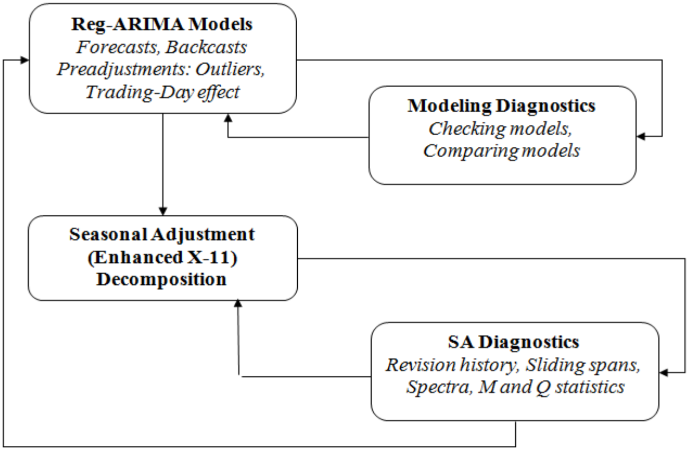
On the Menu today ....

- The RegARIMA model.
- The various kinds of outliers and ruptures.
- Trading-day and Calendar Effects

- Playing with JD+: Is there any calendar effects in your data?

X-13ARIMA-SEATS

Two building blocks: RegARIMA and X11.



The RegARIMA model

A usual additive model for SA is:

$$X_t = TC_t + S_t + O_t + TD_t + MH_t + I_t,$$

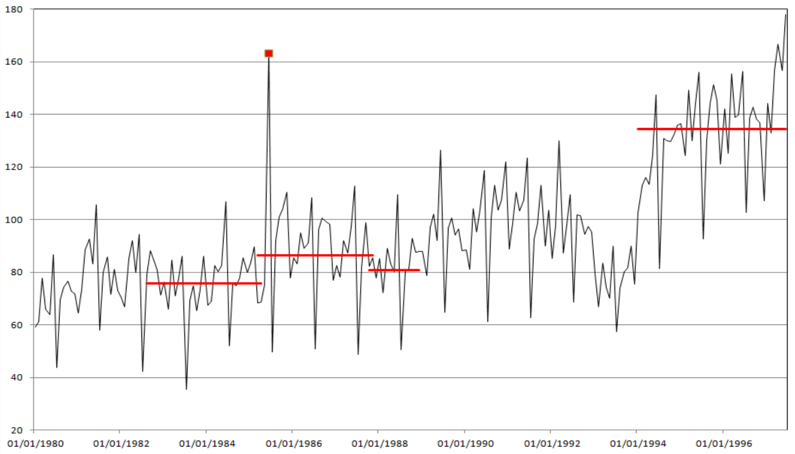
Where all the components are orthogonal and are non-seasonal (except S_t of course).

A SA in 2 steps:

Step 1: $X_t = O_t + TD_t + MH_t + Z_t,$ where Z_t follows an ARIMA model.
Step 2: Decomposition of $Z_t = TC_t + S_t + I_t$

A First Example

Various ruptures in the series.



Additive Outliers

Additive outliers (AO) represent a one-off peak or trough in the time series at a single observation are defined as:

$$AO_t = \begin{cases} 1 & \text{if } t = t_0 \\ 0 & \text{if } t \neq t_0 \end{cases} \tag{1}$$

Temporary Changes

Temporary changes (TC) where the change in the level of the time series is not of permanent but temporary nature, i.e. the effect decreases exponentially ($0 < \alpha < 1$) with the course of time, have the following representation:

$$TC_t = \begin{cases} 0 & \text{if } t < t_0 \\ \alpha^{t-t_0} & \text{if } t \geq t_0 \end{cases} \tag{2}$$

Level Shifts

Level shifts (LS) permanently increase or decrease the (transformed) data by some constant factor prior to a certain observation can be written as:

$$LS_t = \begin{cases} -1 & \text{if } t < t_0 \\ 0 & \text{if } t \geq t_0 \end{cases} \tag{3}$$

Ramps

Ramps (RP) allow for a smooth, linear transition between two time points (t_0 is the start date and t_1 the end date) unlike the abrupt change associated with level shifts are given by:

$$RP_t = \begin{cases} -1 & \text{if } t \leq t_0 \\ (t - t_0)/(t_1 - t_0) - 1 & \text{if } t_0 < t < t_1 \\ 0 & \text{if } t \geq t_1 \end{cases} \tag{4}$$

Temporary Level Shifts

Temporary level shifts (TL) where the level shift is of temporary rather than permanent nature it follows that:

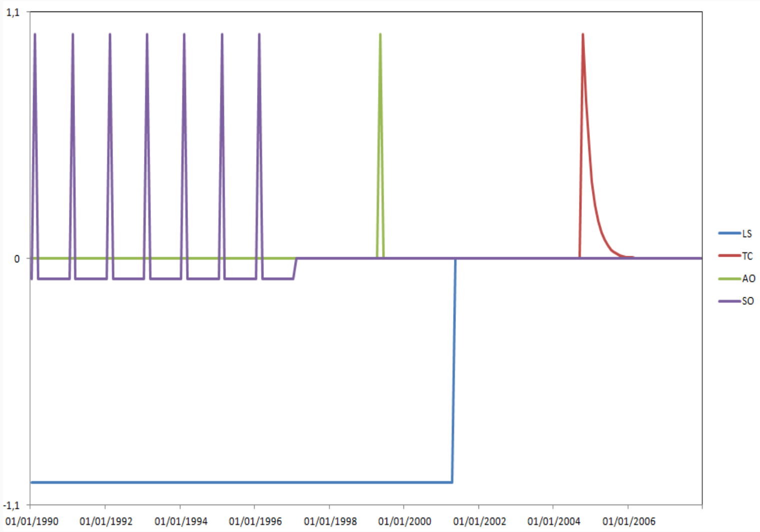
$$TL_t = \begin{cases} 0 & \text{if } t \leq t_0 \\ 1 & \text{if } t_0 < t \leq t_1 \\ 0 & \text{if } t > t_1 \end{cases} \tag{5}$$

Seasonal Outliers

Seasonal outliers (SO) do not affect the level of the time series but the seasonal pattern by allowing for an abrupt increase or decrease of the seasonal component for a specific month ($s = 12$) or quarter ($s = 4$), hence a permanent seasonal break, of which the formula is:

$$SO_t = \begin{cases} 0 & \text{if } t \geq t_0 \\ -1 & \text{if } t < t_0 \text{ and } t \text{ same month/quarter as } t_0 \\ 1/(s-1) & \text{otherwise} \end{cases} \tag{6}$$

Regressors



1,1
0
-1,1
01/01/1990
01/01/1992
01/01/1994
01/01/1996
01/01/1998
01/01/2000
01/01/2002
01/01/2004
01/01/2006
LS
TC
AO
SO

Characteristics of different types of outliers

Type of outlier	Component	Durability of impact	Visible in SA data?	Frequency
Additive outlier	Irregular	Temporary	Yes	Common
Temporary change	Irregular	Temporary	Yes	Rare
Level shift	Trend-cycle	Permanent	Yes	Common
Ramp	Trend-cycle	Permanent	Yes	Rare
Temporary level shift	Trend-cycle	Temporary	Yes	Rare
Seasonal outlier	Seasonal	Permanent	No	Sometimes

Calendars

Calendars

Calendars are normally based on astronomical events; and the 2 most important astronomical objects are the Sun and the Moon.

- Solar calendars (Gregorian, Julian) are based on the motion of the Earth around the Sun. The calendar year approximates the "tropical year"; 12 months strongly linked to the seasons (Seasonal effect).
- Lunar calendars (Islamic) are based on the motion of the Moon around the Earth. A month approximates the "synodic month", the time between 2 "new moons".
- Solar/lunar calendars (Chinese, Hebrew, Hindu) try to achieve the 2 approximations.

The Various Calendar Effects

Months and Quarters

Most of economic series are recorded "according to the Gregorian calendar" on a monthly or quarterly basis .....

- But months (quarters) are not really comparable:
  - ☐ Not the same number of days
  - ☐ Not the same number of Mondays, Tuesdays .....
  - ☐ National public holidays
- And this may have an impact on the observed variable.
  - ☐ Retail sales, Transportation etc.
- Important temporal and spatial differences

Temporal differences

Big differences if we take into account the National specificities.
Example: Number of working days (Monday to Friday) in France by quarter.

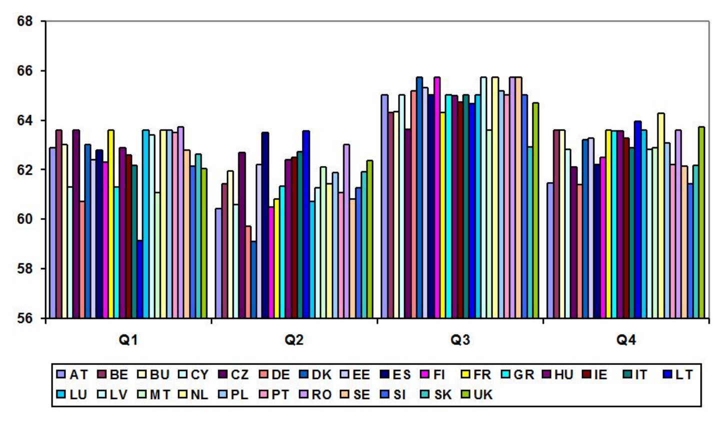
- Compare 2017Q2 to 2016Q2 (annual growth rate)
- Compare 2017Q2 to 2017Q1 (quarterly growth rate)

	Gregorian calendar					National calendar				
Year	Q1	Q2	Q3	Q4	Total	Q1	Q2	Q3	Q4	Total
2016	65	65	66	65	261	63	63	64	63	253
2017	65	65	65	65	260	65	60	63	63	251
2018	65	65	65	66	261	64	60	64	64	252
2019	64	65	66	66	261	63	61	65	64	253
2020	65	65	66	65	261	64	63	64	63	254
2021	65	65	65	65	260	65	61	63	63	252
2022	65	65	65	66	261	64	61	64	64	253

Geographical differences

Number of working-days (Monday to Friday) by quarter for EU countries.

A very busy graph ⇔ large spatial differences.



The various calendar effects

- The seasonality (due to the Gregorian solar calendar)
- Length-of-Month (LOM) effect
- "Day of the week" (Trading-day) effect
    - Ex.: Retail trade turnover is likely to be more important on Saturdays than on other days.
- Public Holidays
    - Most of statutory holidays are linked to a date, not to a day of the week (Christmas).
- Moving Holidays
    - Some holidays "move" across the year (Easter, Ramadan, Chinese New Year, Diwali etc.).
    - Usually occurs when you mixed 2 calendars.

Detecting "pure" Trading-Day Effects

Detecting "pure" Trading-Day Effects

Hopefully, the Gregorian calendar is quite regular:

- The structure of the week is periodic with period 7;
- The length of the year is of 365 or 366 days. Leap Year?
  - (MOD(year,4)=0 AND MOD(year,100) NE 0) OR MOD(year,400)=0.
  - 97 leap years in a 400-year cycle.
  - The length of the year is periodic with period 400 years.
- Integer number of weeks in a complete cycle (20871);
- Thus the Gregorian calendar is periodical with period 400 years.
- Detection of a trading-day effect:
  - Using a Reg-ARIMA model.

Did you know that .....

2 pages from a Swedish almanac from 1712.

- Top left corner: "Februarius", At the top of the right page: XXX
- And at the bottom weather forecasts !!! (Snöö = snow)

Februarius Göijemånadt

Sool up 7/37. ned. 4/23. Dag 9/28 natt. 14/32.

Gamal Styl. Nya St.

1 d Brigita 11

2 e Kyndelsmässe wackert 12

3 f Blasius Winter. 13

Om Hwetet och Ogräset/ Matth. 13/24.

4 G S.e.Trett. 7/24.e. 14

5 a Agatha Wäder/ 15

6 b Dorothea 16

7 c Richardus Snöö 17

8 d Helena 18

9 e 40 Martyrer 11/46.f. 19

10 f Apollonius Töö 20

Om the 10 Jungfrur Matth. 25. 1.

11 G 6 S.e.Tr. Disiinge 21

12 a Eulalia 12/28 e. 22

13 b Nicephorus Wäder/ 23

haar XXX Dagar.

Sool up 6/39. ned. 5/30. Dag. 10/32. tim. N. 13/28.

14 c Valentin 24

15 d Sigfrid 25

16 e Juliana Frost/ 26

17 f Casparus 27

Om Arbetarna i Wingården/ Matth. 20/1.

18 G Septuag. Concordia 28

19 a Susanna 12/58. e. 29

20 b Joachim 1 Mart.

21 c Benedictus Blåst/ 2

22 d Victor ohrwäder/ 3

23 e Tertullianus 4

24 f Matthias Kalt/ 5

Om fyra handa Sädes åker. Luc. 8/4.

25 G Sexages. Skottdagen. 6

26 a Victoria 12/7. e. 7

27 a Leander 8

28 b Romanus ostadigt 9

29 c Nestor 10

30 c Tillökad: Snöö/ 11

Moving Holidays

Moving Holidays

- Basically due to the use of 2 different calendars at the same time
- Christian Easter
  - Easter Sunday is the first Sunday after the first full moon after vernal equinox.
  - Linked to the astronomic calendar
- Ramadan, Aid-el-Fitr, Aid-el-Adha etc.
  - Linked to the Islamic calendar (lunar)
- Chinese New Year, Diwali, Pesach etc.
  - Lunar/solar calendars

Example: Calculating Islamic Festival Dates

- One year has 354 or 355 days; 12 months with 29 or 30 days;
- In "Leap Years", the last month (Joumada al Oula) counts 30 days;
- The years follow a 30-year cycle;
- Several "rules" for the "leap years":
  - ☐ 2, 5, 7, 10, 13, 15, 18, 21, 24, 26, et 29 (Kuwait algorithm)
  - ☐ 2, 5, 7, 10, 13, 16, 18, 21, 24, 26, et 29 (common version)
  - ☐ 2, 5, 8, 10, 13, 16, 19, 21, 24, 27, et 29 (Indian tables)
  - ☐ 2, 5, 8, 11, 13, 16, 19, 21, 24, 27, et 30
- On this 30-year cycle the year average length is very close to the length of the lunar year (<35s).
- Ras-El-Am = MDY(1,25,1583) and you follow the cycle
- But you might have to adapt this calendar to the real observation of the Moon.

Building a National Calendar

National Calendars

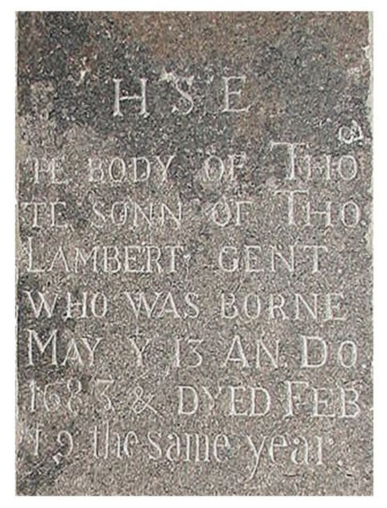
- Building a National calendar is not easy, it could even become a nightmare very quickly.
- A calendar is alive:
  - ☐ Public or bank holidays appear and disappear;
  - ☐ Substitute days (UK, Canada ....);
  - ☐ Special days (UK: Queen's Jubilee ....)
- Exception is ..... the Rule!
  - ☐ Provincial specificities (Germany, Canada, India ....) etc.
- How to construct a European calendar?
  - ☐ Weighting problem
- Is it really worth it? If it is complex, keep it simple!

The Maldivian Calendar

We focus on the following 12 holidays:

- 5 fixed holidays in the Gregorian calendar: New Year (January 1), Labor Day (May 1), Independence Day (July 26, 2 days), Victory Day (November 3), Republic Day (November 11);
- 7 Islamic moving holidays: Islamic New Year, Prophet Muhammad's Birthday, The Day Maldives Embraced Islam, First of Ramazan, Eid-ul Fithr (3 days), Hajj Day and Eid-ul Al'haa (4 days).

Did you know that ......



- The tombstone tells about a boy who was born on 13 May 1683 and died on 19 February of the same year;

- In 1683 in England, new year's day was 25 March;
- Indeed, would not be Spring a much better date to begin a new year?

Modeling the Trading-Day Effect

A Basic Regression Model

A usual additive model for SA is:

$$X_t = TC_t + S_t + O_t + TD_t + MH_t + I_t,$$

Where all the components are orthogonal and are non-seasonal (except S_t of course).

A basic model for trading-day could be:

$$X_t = \sum_{i=1}^{i=7} \alpha_i N_{it} + \epsilon_t,$$

Where:

- N_{it} = # of Mondays (i=1), ..., Sundays (i=7) in month t.
- α_i is the effect of day i. (constant on the period)
- ϵ_t is a white noise (X11, X12) or an ARIMA model (X12, TS)

Improving the Model

The (Reg-ARIMA) model presents 3 problems:

- The regressor covariance matrix is ill-conditioned and its inverse is unstable $\Longrightarrow$ estimates are also unstable.
- TD and seasonality are not independent
- We also would like that the TD effect is equal to 0 over a week

To solve most of these problems, the most popular model (X12, TS) uses contrasts and do the estimation on the raw data:

$$X_t = \beta_0 LY_t + \sum_{j=1}^{j=6} \beta_j (N_{jt} - N_{7t}) + \epsilon_t \quad \text{and} \quad \widehat{\beta}_7 = -\sum_{j=1}^{j=6} \widehat{\beta}_j$$

Stability of the Estimates

The inverse covariance matrix of the contrasts is more stable.

Ratio[min(λ)/ max(λ)], computed over the 400-year cycle; all series of length 5 to 25 years.

Length	Gregorian Calendar			French National Calendar	
(in years)	7 regressors	6 contrasts	2 contrasts	6 contrasts	2 contrasts
5	0.0005	0.0321	0.0243	0.1773	0.1444
10	0.0005	0.0329	0.0243	0.1935	0.1442
15	0.0005	0.0331	0.0243	0.1975	0.1440
20	0.0005	0.0331	0.0243	0.1997	0.1437
All	0.0005	0.0328	0.0243	0.1919	0.1441

The ratio is multiplied by about 6 which shows a much better stability.

Seasonality

Contrasts remove a large part of TD seasonality:

- The seasonal part is the average number of a specific day over the 400-year cycle;
- The averages (see Table) are quite close and taking the difference reduces the seasonality.

Month	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	4.4300	4.4250	4.4300	4.4275	4.4300	4.4300	4.4275
2	4.0325	4.0375	4.0325	4.0375	4.0325	4.0350	4.0350
3	4.4300	4.4275	4.4300	4.4250	4.4300	4.4275	4.4300
4	4.2850	4.2875	4.2850	4.2875	4.2850	4.2850	4.2850
5	4.4300	4.4250	4.4300	4.4275	4.4300	4.4300	4.4275
6	4.2850	4.2875	4.2850	4.2850	4.2850	4.2850	4.2875
7	4.4300	4.4275	4.4300	4.4300	4.4275	4.4300	4.4250
8	4.4275	4.4300	4.4250	4.4300	4.4275	4.4300	4.4300
9	4.2875	4.2850	4.2875	4.2850	4.2850	4.2850	4.2850
10	4.4250	4.4300	4.4275	4.4300	4.4300	4.4275	4.4300
11	4.2875	4.2850	4.2850	4.2850	4.2850	4.2875	4.2850
12	4.4275	4.4300	4.4300	4.4275	4.4300	4.4250	4.4300

Usual Models

- The 7-regressor model:

$$X_t = \beta_0 LY_t + \sum_{j=1}^{j=6} \beta_j (N_{jt} - N_{7t}) + \epsilon_t \quad \text{and} \quad \widehat{\beta}_7 = -\sum_{j=1}^{j=6} \widehat{\beta}_j$$

- The 2-regressor model (weekdays vs. weekend)

$$\beta_6 = \beta_7 \quad \text{and} \quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$$
$$\textstyle\sum_{j=1}^{j=7} \beta_j = 0 \Rightarrow \beta_6 = \beta_7 = -\frac{5}{2}\beta_1$$

$$X_t = \beta_0 LY_t + \beta_1 \left[\textstyle\sum_{j=1}^{j=5} N_{jt} - \frac{5}{2}(N_{6t} + N_{7t}) \right] + \epsilon_t$$

Limits of these (too) simple models

- Do not take into account National specificities;
- 2 sets of regressors only : 1 or 6 contrasts. Reality might be more complex (Saturday for Retail Trade)
- Effects are constant over time. But things and laws change (Retail Trade and Sundays)
- More flexibility is needed to construct and adapt regressors.
  - Groupings;
  - Take then care of potential seasonality.

Taking into account the National calendar

- Complete Model with 15 variables: distinction between Monday off and Monday in.

$$TD_t = \beta_0 LY_t + \sum_{j=1}^{j=14} \beta_j N_{jt}$$

- Where N_{jt} are the # of Mondays in, Tuesdays in, ..., Sundays in, Mondays off, Tuesdays off, ..., Sundays off.
- But not very parsimonious: hypothesis on the various days.

Example

- All days off are equivalent to a Sunday.

$$\beta_7 = \beta_8 = \beta_9 = \cdots = \beta_{13} = \beta_{14}$$

- And we obtain the following model:

$$TD_t = \beta_0 LY_t + \sum_{j=1}^{j=6} \beta_j \left[N_{jt} - \frac{1}{8} \left(\sum_{j=7}^{j=14} N_{jt} \right) \right]$$

- But contrasts do not remove the seasonality anymore: Need to seasonally adjust the regressors using the 400-year periodicity of the calendar or the periodicity of the moving holiday (Easter, Ramadan, Chinese New Year, Diwali, Pesach etc.).

Modeling moving holidays

Modeling moving holidays

Moving Holidays might have 2 different effects:

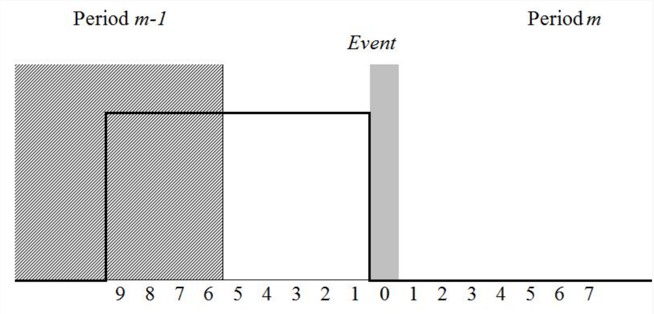
- Days off are often associated to these holidays to celebrate the event. This effect is accounted for in the Trading-Day effect;
- They might have an impact on the economy. For examples, members of a family might travel to celebrate the event together, the prices of some commodities might increase before the event etc.

This is the second effect we want to capture here using an “impact model”.

An Example: Easter

- We will take the example of Easter; very similar to Pesach, Diwali in principle.
- Models for this holiday are already implemented in JDemetra+:
  - Several models in X-11 and X-11-ARIMA (irregular);
  - Two strategies in X-13ARIMA-SEATS: looking for in the raw data or in the irregular component;
  - Two possible models: SCeaster[w] and Easter[w]; very similar models.
- Easter[w]:
  - The activity is affected the w days before Easter Sunday;
  - The activity level remains constant on the w-day period.
  - Easter can move from March 22 and April 25. Therefore, the 2 months can be impacted according to the year.

The Easter[w] Model



- If i denotes the year, j the month. Let us note $1 <= w <= 25$ and n_{ij} the number of the w days falling in month j of year i.
- The regressor associated to this model is: $X_{ij} = n_{ij}/w$.
- Should be seasonally adjusted (eastermeans). Then very close to SCeaster[w].

More General Models