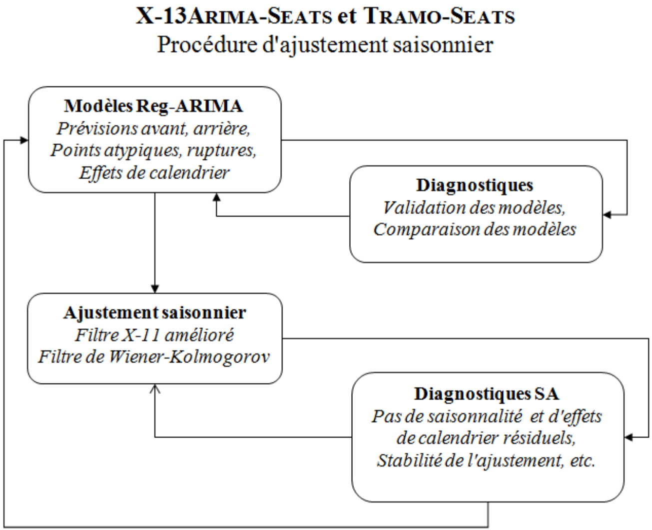
SACE Meeting #5

# Stability of Reg-ARIMA estimates: the case of leap year and outliers

Dominique Ladiray et Alain Quartier-la-Tente
Insee, Seasonal Adjustment Centre of Excellence (SACE)

# Introduction to the seasonal adjustment procedure



# Mathematical Writing of Reg-ARIMA

Mathematical writing of the Reg-ARIMA Model in Seasonal Adjustment:

$$\left.\begin{array}{ll}\text{Additive:} & Y_t \\ \text{Multiplicative:} & \log(Y_t)\end{array}\right\} = \underbrace{\beta_0 LY_t + \beta_1 WD_t}_{\text{WD regressors}} + \underbrace{\sum_i \gamma_i O_{i,t}}_{\text{outliers}} + \underbrace{\varepsilon_t}_{\sim ARIMA}$$

# Mathematical Writing of Reg-ARIMA

Mathematical writing of the Reg-ARIMA Model in Seasonal Adjustment:

$$\left.\begin{array}{ll}\text{Additive:} & Y_t \\ \text{Multiplicative:} & \log(Y_t)\end{array}\right\} = \underbrace{\beta_0 LY_t + \beta_1 WD_t}_{\text{WD regressors}} + \underbrace{\sum_i \gamma_i O_{i,t}}_{\text{outliers}} + \underbrace{\varepsilon_t}_{\sim ARIMA}$$

The goal of the study: illustrate some problems of instability of the estimates with examples on:

- the leap year adjustment
- the outliers estimates
- the identification of the ARIMA model

# Sommaire

# Quand faut-il le corriger ?

A leap year: one additional day in February $\simeq$ every 4 years

$\rightarrow$ takes into account the "length of the month" effect: it is a calendar effect

When to correct it ?

# Quand faut-il le corriger ?

A leap year: one additional day in February $\simeq$ every 4 years

$\rightarrow$ takes into account the "length of the month" effect: it is a calendar effect

When to correct it ?

According to the guidelines on seasonal adjustment, do so when:

- there is an economic sense to doing it
- the effect is stable and statistically significant

# Quand faut-il le corriger ?

A leap year: one additional day in February $\simeq$ every 4 years

$\rightarrow$ takes into account the “length of the month” effect: it is a calendar effect

When to correct it ?

According to the guidelines on seasonal adjustment, do so when:

- there is an economic sense to doing it
- the effect is stable and statistically significant

Study of European IPI (1330 series): the leap year effect exists (but not always measurable due to collection)

# To methods to proceed to the adjustment

1. With the Reg-ARIMA model:

$$LY_t = \begin{cases} 0.75 & \text{in February during leap years} \\ -0.25 & \text{in February during non leap years} \\ 0 & \text{for all other months} \end{cases}$$

# To methods to proceed to the adjustment

1. With the Reg-ARIMA model:

$$LY_t = \begin{cases} 0.75 & \text{in February during leap years} \\ -0.25 & \text{in February during non leap years} \\ 0 & \text{for all other months} \end{cases}$$

2. Correcting values prior to modelling multiplying by a fixed proportion:

$$\begin{cases} \frac{28.25}{29} \simeq 0.974 & \text{in February during leap years} \\ \frac{28.25}{28} \simeq 1.009 & \text{in February during non leap years} \\ 1 & \text{for all other months} \end{cases}$$

# To methods to proceed to the adjustment

1. With the Reg-ARIMA model:

$$LY_t = \begin{cases} 0.75 & \text{in February during leap years} \\ -0.25 & \text{in February during non leap years} \\ 0 & \text{for all other months} \end{cases}$$

2. Correcting values prior to modelling multiplying by a fixed proportion:

$$\begin{cases} \frac{28.25}{29} \simeq 0.974 & \text{in February during leap years} \\ \frac{28.25}{28} \simeq 1.009 & \text{in February during non leap years} \\ 1 & \text{for all other months} \end{cases}$$

# To methods to proceed to the adjustment

1. With the Reg-ARIMA model:

$$LY_t = \begin{cases} 0.75 & \text{in February during leap years} \\ -0.25 & \text{in February during non leap years} \\ 0 & \text{for all other months} \end{cases}$$

2. Correcting values prior to modelling multiplying by a fixed proportion:

$$\begin{cases} \frac{28.25}{29} \simeq 0.974 & \text{in February during leap years} \\ \frac{28.25}{28} \simeq 1.009 & \text{in February during non leap years} \\ 1 & \text{for all other months} \end{cases}$$

$\rightarrow$ Study of estimates of the 1st method

# Methodology used

Methodology: **identified** the model throughout the entire sample (ARIMA, outliers, etc.) and re-estimate month-by-month the past values **setting** the first estimated date

# Methodology used

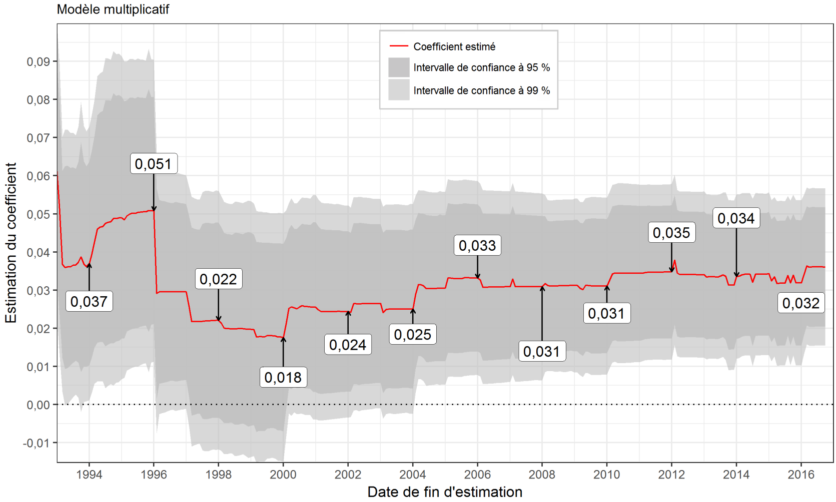
Methodology: **identified** the model throughout the entire sample (ARIMA, outliers, etc.) and re-estimate month-by-month the past values **setting** the first estimated date

We considerded that the estimate has converged whe the estimated coefficient remains:

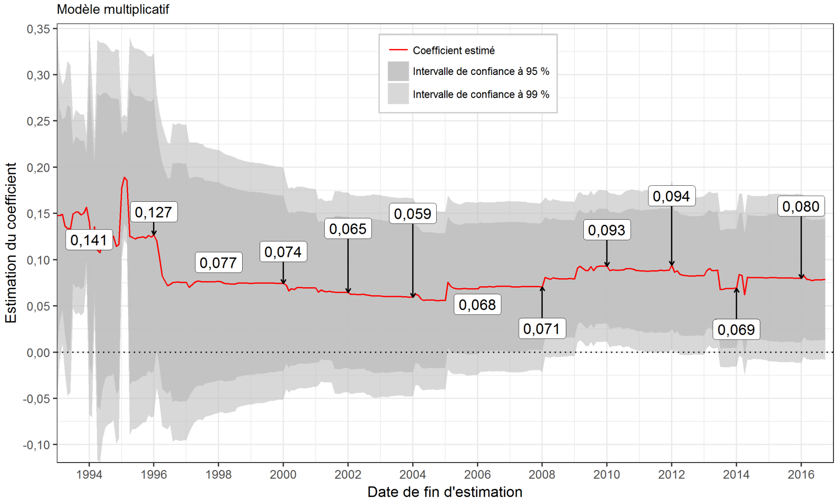
- positive
- not significantly different from last estimation
- significant: stability of the choice to adjust the leap year effect

$\Longrightarrow$ European IPI: 410 converge

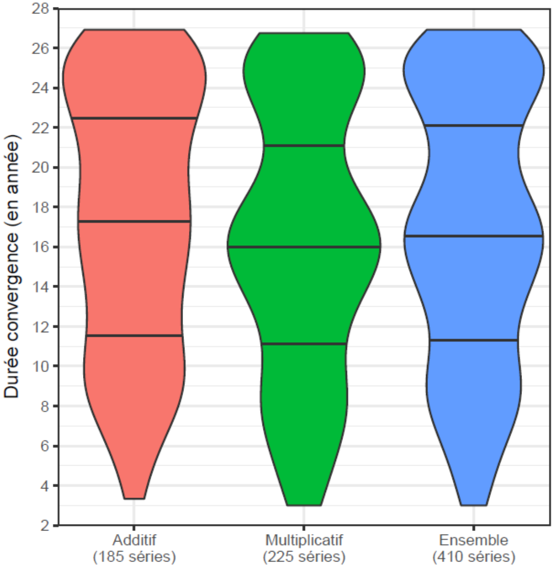
# Examples (1/2): IPI FR-0610 (extraction of crude petroleum)



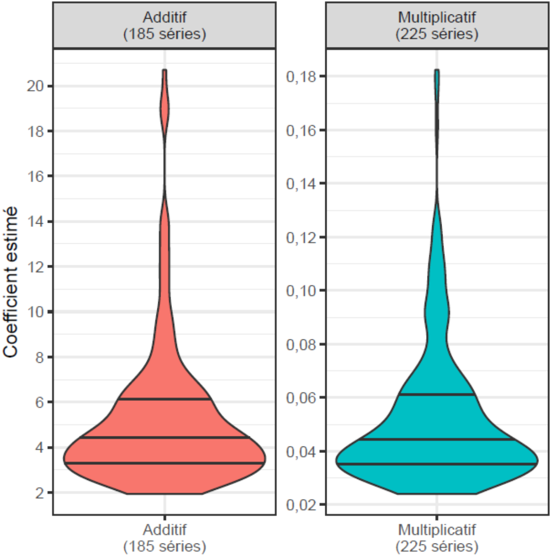
# Examples (2/2): IPI FR-1391 (manufacture of knitted and crocheted fabrics)



# A rather slow convergence. . .



# ... Towards a value not always coherent



# Comparison of the two correction methods

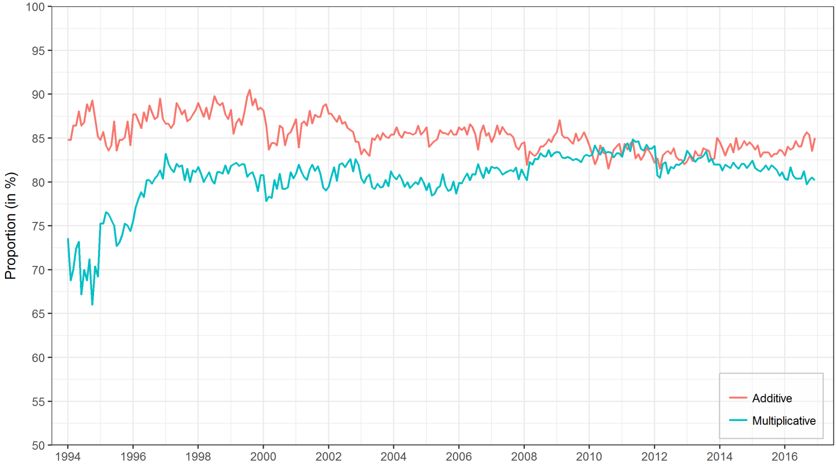
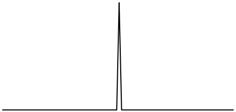



Figure 1: Percentage of series for which the AICC of the 2nd method (LY pre-adjustment) is lower than the AICC of the 1st method (LY regressor)

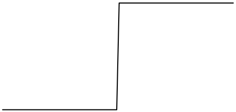
# Sommaire

# Usuals outliers

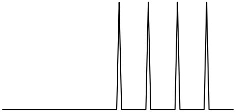
**Additive outlier** (AO)



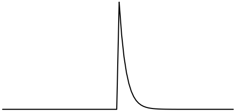
**Level Shift** (LS)



**Seasonal Outlier** (SO)



**Transitory Change** (TC)



# Methodology used

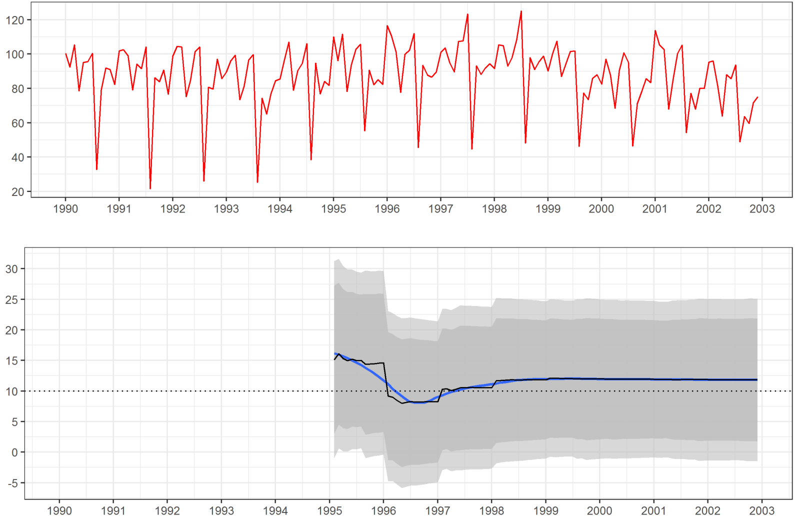
On European IPI: 1. **identification** and **estimation** of the model over 12 years

2. The raw series is corrected for outliers in the year of the introduction of the break and it is rebased to 100 at the month of the introduction of the break
3. simulation of a break, 5 years after the beginning of the time series date, of level 10 for a **additive** model
4. The regressor related to the simulated break is added to the Reg-ARIMA model and its coefficient it is estimated by freezing the estimates of all other parameters and **setting** the first estimated date

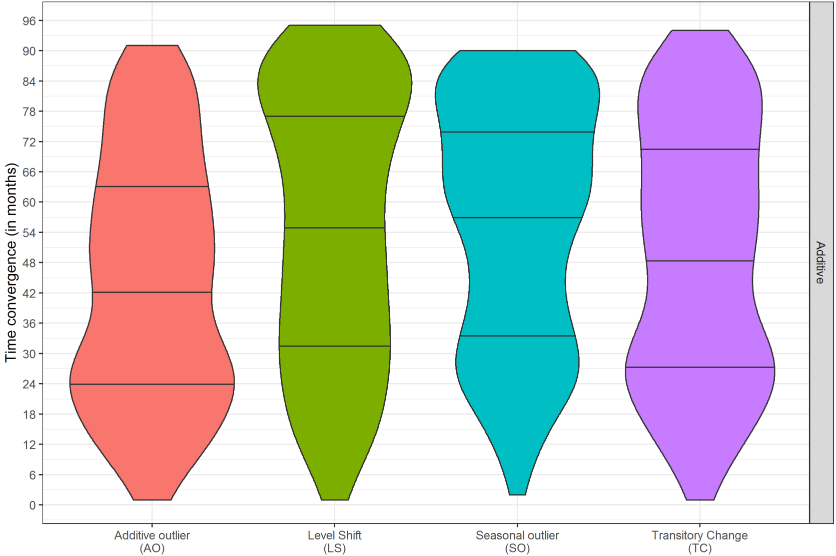
We considered that the series has converged when:

$$\left|\frac{\text{estimated value}}{\text{last estimated value}} - 1\right| < 5\,\%$$

# Example of a AO for IPI IT-1413 (manufacture of other outerwear)



# A rather slow convergence..



# . . . But not always to the right value

| | Minimum | 25 % | 50 % | 75 % | Maximum |
|---|---|---|---|---|---|
| **Additive models** | | | | | |
| Additive outlier (AO) | -11.6 | 7.8 | 11.1 | 14.2 | 36.9 |
| Level Shift (LS) | -11.4 | 5.6 | 9.3 | 12.7 | 49.8 |
| Seasonal outlier (SO) | -5.8 | 7.3 | 8.8 | 11.0 | 31.1 |
| Transitory Change (TC) | -17.4 | 6.5 | 10.2 | 14.1 | 47.2 |

# Sommaire

# Identification of two equivalent models

We use the same model in two different forms mathematically equivalent:

1. The leap year regressor is added as a working days regressor
2. The leap year regressor is added as an external regressos

$\rightarrow$ study of the automatic models

# Different automatic models

| Le régresseur LY est dans les effets de calendrier | Le régresseur LY est dans les régresseurs externes |
|---|---|

**Le régresseur LY est dans les effets de calendrier**

**Summary**

Estimation span: [1-1990 - 11-2016]
323 observations
Trading days effects (7 variables)
3 detected outliers

**Arima model**

[(2,0,0)(0,1,1)]

| | Coefficients | T-Stat | P[\|T\| > t] |
|---|---|---|---|
| Phi(1) | -0,5256 | -9,46 | 0,0000 |
| Phi(2) | -0,2878 | -5,17 | 0,0000 |
| BTheta(1) | -0,7913 | -20,56 | 0,0000 |

Correlation of the estimates

| | Phi(1) | Phi(2) | BTheta(1) |
|---|---|---|---|
| Phi(1) | 1,0000 | -0,7388 | -0,0184 |
| Phi(2) | -0,7388 | 1,0000 | 0,0489 |
| BTheta(1) | -0,0184 | 0,0489 | 1,0000 |

Leap year

| | Coefficients | T-Stat | P[\|T\| > t] |
|---|---|---|---|
| Leap year | 4,3861 | 2,65 | 0,0085 |

**Le régresseur LY est dans les régresseurs externes**

**Summary**

Estimation span: [1-1990 - 11-2016]
323 observations
No trading days effects
8 detected outliers

**Arima model**

[(0,1,1)(0,1,1)]

| | Coefficients | T-Stat | P[\|T\| > t] |
|---|---|---|---|
| Theta(1) | -0,5051 | -10,08 | 0,0000 |
| BTheta(1) | -0,7533 | -18,80 | 0,0000 |

Correlation of the estimates

| | Theta(1) | BTheta(1) |
|---|---|---|
| Theta(1) | 1,0000 | 0,0280 |
| BTheta(1) | 0,0280 | 1,0000 |

User variables

| | Coefficients | T-Stat | P[\|T\| > t] |
|---|---|---|---|
| Leap year | 4,5569 | 2,92 | 0,0038 |

# Sommaire

# CConclusion and recommendations (1/2)

Simulations are **questionables** and **can be improved** but highlight a **potential instability** of Reg-ARIMA models often used as black boxes

# CConclusion and recommendations (1/2)

Simulations are questionables and can be improved but highlight a potential instability of Reg-ARIMA models often used as black boxes

Instabilities have a **limited effect** on the SA-WD series... but have an impact on the **short term** history and on **revisions** !

# CConclusion and recommendations (1/2)

Simulations are questionables and can be improved but highlight a potential instability of Reg-ARIMA models often used as black boxes

Instabilities have a limited effect on the SA-WD series... but have an impact on the short term history and on revisions !

Automatic algorithms use in X-13ARIMA-SEATS and TRAMO-SEATS are **important** and very **useful**

# Conclusion and recommendations (2/2)

Specify the model **beforehand** at the level of the series **series**:

- base selection procedures on economic reasoning (pay attention to long series)
- do not use methods like black boxes...

# Conclusion and recommendations (2/2)

Specify the model **beforehand** at the level of the series **series**:

- base selection procedures on economic reasoning (pay attention to long series)
- do not use methods like black boxes... Otherwise, you will be like this statistician who...

# Thank you for your attention

*"He uses statistics as a drunk man uses lamp-posts: for support rather than for illumination.*

Quote widely attributed to Andrew Lang (1844-1912)