European Commission

Time series

Short tutorial

ESTP training

15-17/10/2024

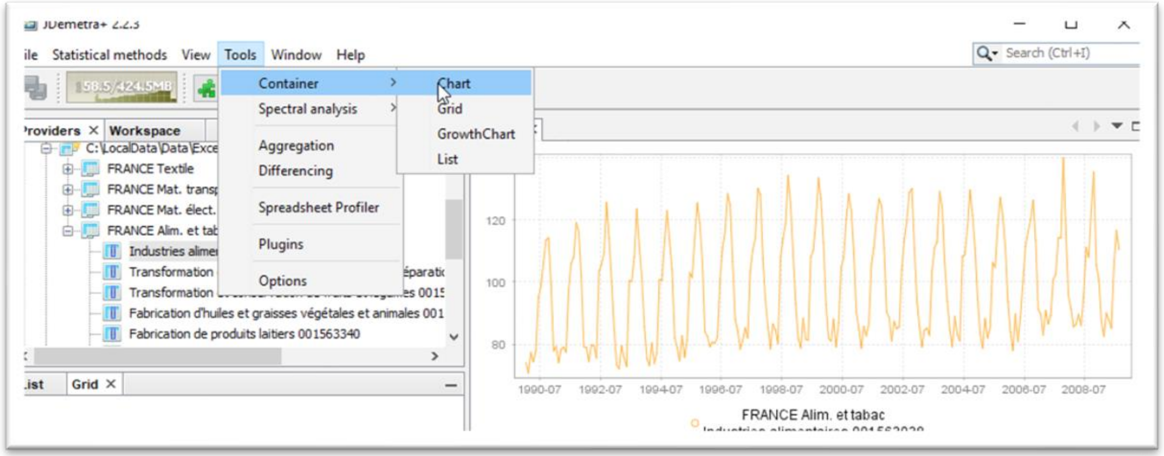
1. Outline

- “Regular” time series
- Basic tools
  - Auto-correlation function
  - Periodogram
  - Linear filter
- Stationarity
- Arima model

2. Time series

- We will only consider « regular » time series
  - Monthly, quarterly…[, yearly] time series
  - Annual frequency: 12, 6, 4, 3, 2, 1

- [JD+ 3.0 deal with more general time series]



3. Basic tools

- Auto-correlation function and auto-covariance function (ACF)
  - Measurement of “patterns” in time series (x_t, $_{0 \leq t < n}$)
  - $acf(k) = \rho_k = \frac{\sum_{t=k}^{n-1}(x_t - \bar{x})(x_{t-k} - \bar{x})}{[\sum_{t=0}^{n-1}(x_t - \bar{x})^2]}$
  - If we assume that $\bar{x} = 0$, $\rho_k = \frac{\sum_{t=k}^{n-1} x_t x_{t-k}}{[\sum_{t=0}^{n-1} x_t{}^2]}$

3. Basic tools (cont.)

- Periodogram (see JD+_Seasonality_final_version.docx, pages 12-14).
  - Discrete Fourier's transform of the series $\{x_j\}_{0\le j<n}$
    - Fourier frequencies: $\omega_k = \frac{2\pi}{n} j, \quad 0 \le j < n$
    - $\boldsymbol{e}_j = n^{-1/2}(e^{i\omega_j}, e^{i2\omega_j}, \dots, e^{in\omega_j}), \quad \boldsymbol{F_n} = \{\boldsymbol{e_j}\}$ orthonormal basis of $\mathbb{C}^n$ $\quad (e^{i\omega} = cos\omega + isin\omega)$
    - $x = \sum_{0\le j<n} a_j \boldsymbol{e}_j = \sum_{0\le k<n} x_k \boldsymbol{u}_k, \boldsymbol{u}_k = (0, \dots, 0, 1_k, 0, \dots, 0)$
    - $a_j = n^{-1/2} \sum_{k=0}^{k<n} x_k e^{-ik\omega_j} \quad [or \quad a_j = \sum_{k=0}^{k<n} x_k e^{-ik\omega_j}]$
    - $x_k = n^{-1/2} \sum_{j=0}^{j<n} a_j e^{ij\omega_k} \quad [or \quad x_k = n^{-1} \sum_{j=0}^{j<n} a_j e^{ij\omega_k}]$
    - $p_{\omega_k} = |a_k|^2$
  - The periodogram gives the importance of each (Fourier) frequency in the series
  - Code in toolkit/base/core/data/analysis/Periodogram.java

4. Linear filters

- $y_t = \sum_{i=-\infty}^{+\infty} \beta_i x_{t-i}, \{\beta_i\} \equiv linear\ filter$

- Backward, forward operators: $B^k x_t = x_{t-k}, F^k x_t = x_{t+k}$

- $y_t = P(B)x_t, P(B)y_t = x_t, P(B)y_t = Q(B)x_t$
  - $y_t = x_t - x_{t-1}$
  - $y_t - 0.9y_{t-1} = x_t$ or $y_t = x_t + 0.9x_{t-1} + 0.81x_{t-2} + \ldots$
- $y_t = P(B,F)x_t, P(B,F)y_t = x_t, P(B,F)y_t = Q(B,F)x_t$
  - $y_t = {}^1/_3\, x_{t+1} + {}^1/_3\, x_t + {}^1/_3\, x_{t-1}$

4. Linear filters: properties

- Filter:

$$y_t = \sum_{j=-\infty}^{+\infty} \beta_j x_{t-j}$$

- Frequency response:

$$F(\omega) = \sum_{j=-\infty}^{+\infty} \beta_j e^{-i\omega j}$$

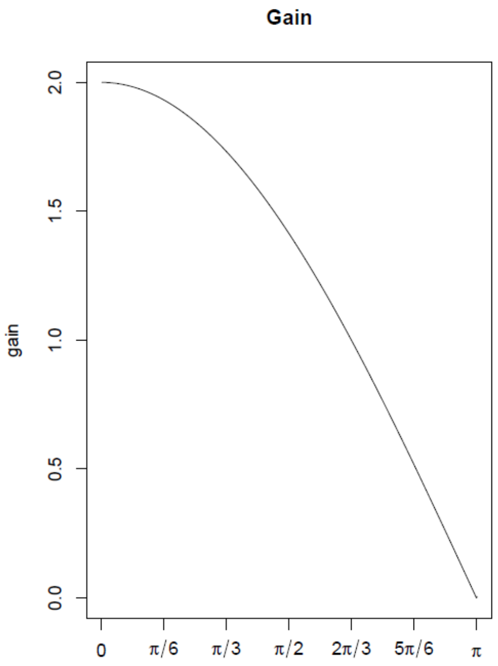
- Gain, phase

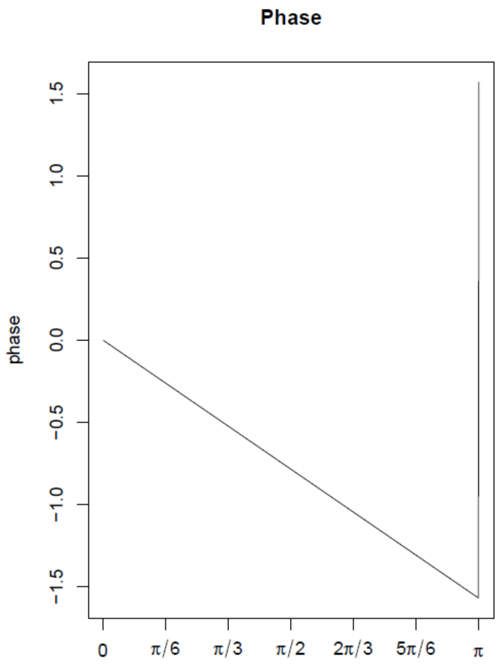
$$|F_t(\omega)|, \arg(F_t(\omega)$$

4. Linear filters: properties

- $y_t = x_t - x_{t-1}$

```r
## Filter x(t)-x(t-1) = (1-B)x(t)

rf<-function(w){
  1+complex(real=cos(w), imaginary = -sin(w))
}

w<-seq(0:600)*(pi/600)
cw<-rf(w)

gain<-Mod(cw)
phase<-Arg(cw)
## or phase<-Arg(cw)/w
```

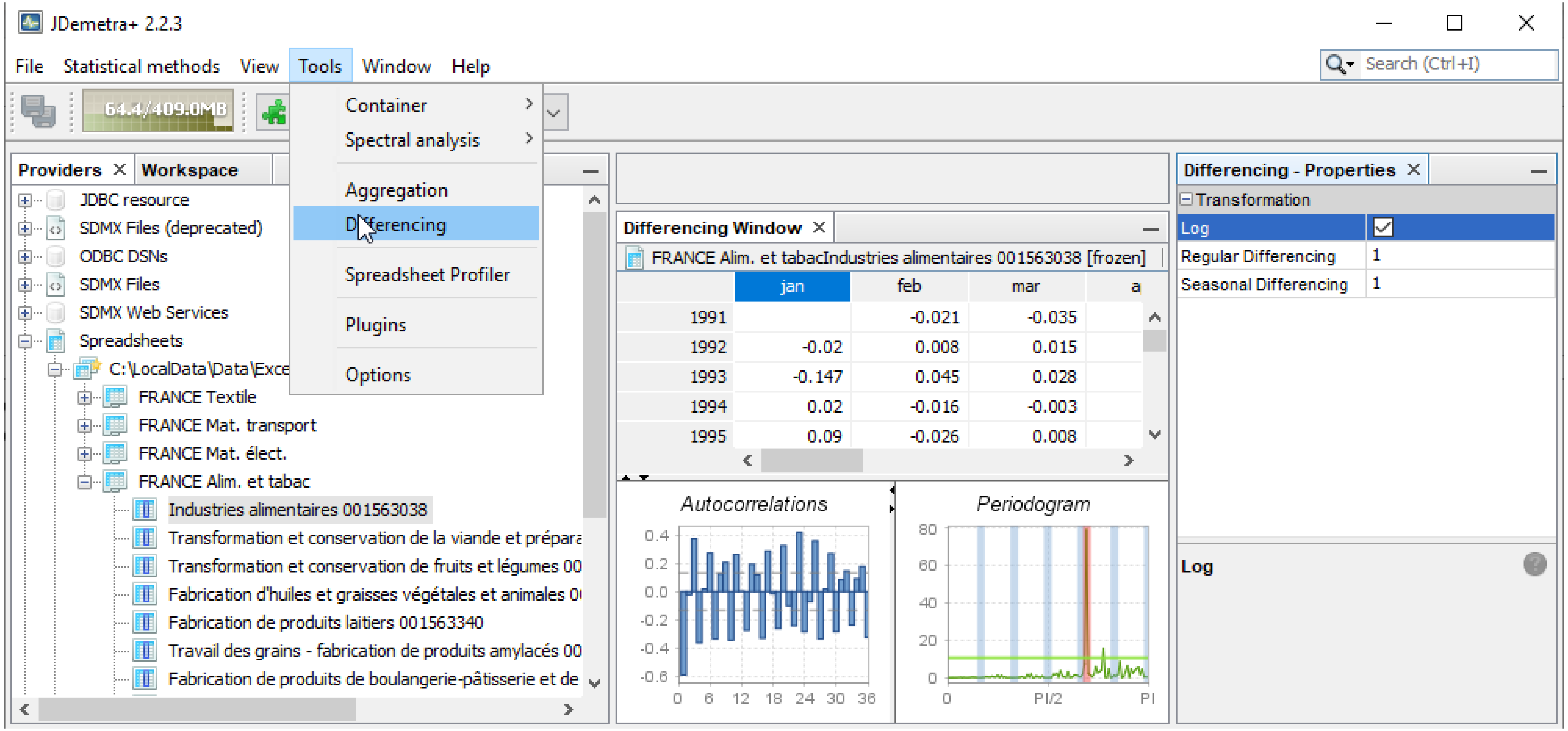




4. Stationarity

- A series is said stationary at the second order, if its mean and variance do not depend on time and if the covariance between lagged series only depends on the difference between lags.
- Many statistical tools/algorithms only apply on stationary time series
- Most economic time series are non-stationary
- (Simple) solutions
  - Log-transformation
  - Differencing
    - $y_t = x_t - x_{t-1}$
    - $y_t = x_t - x_{t-s}$

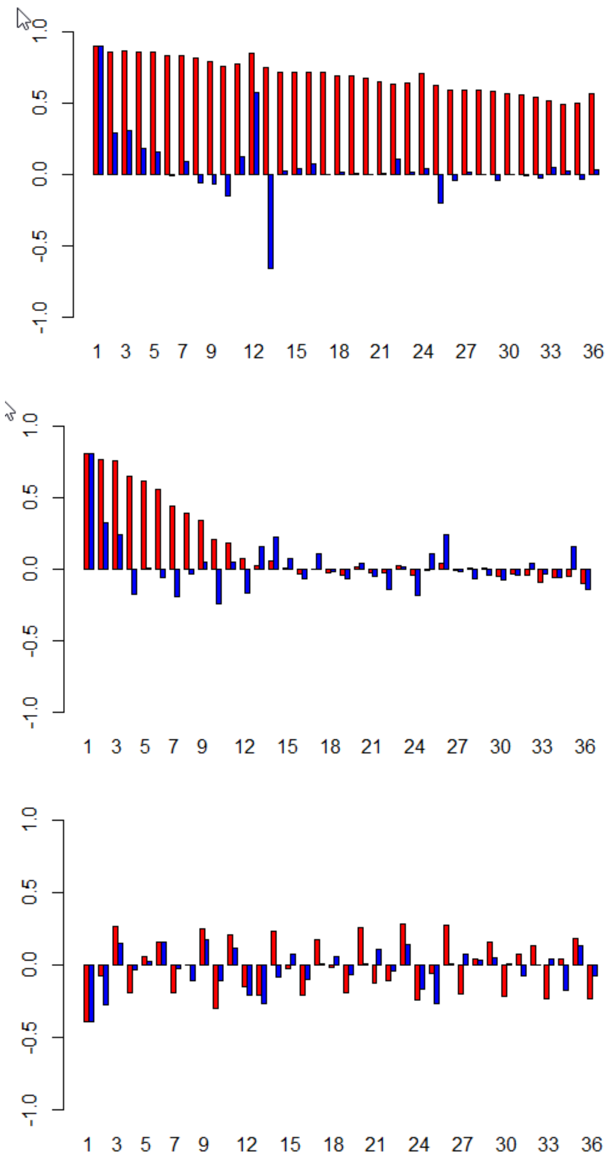
Stationary series



15-17/10/2024

Stationary series in R

```
suppressPackageStartupMessages(library(rjd3modelling))

s<-log(retail$RetailSalesTotal)

plot(s)

ac<-rjd3toolkit::autocorrelations(s, T, n=36)
pac<-rjd3toolkit::autocorrelations.partial (s, T, n=36)

all<-cbind(ac, pac)
barplot(t(all), beside = T, col = c("red", "blue"), names.arg = c(1:36), ylim=c(-1,1))

ds=differences(s,12,F)
#plot(ds, type='l')

ac<-rjd3toolkit::autocorrelations(ds, T, n=36)
pac<-rjd3toolkit::autocorrelations.partial(ds, T, n=36)

all<-cbind(ac, pac)
barplot(t(all), beside = T, col = c("red", "blue"), names.arg = c(1:36), ylim=c(-1,1))

ds=differences(ds,1,F)
#plot(ds, type='l')

ac<-rjd3toolkit::autocorrelations(ds, T, n=36)
pac<-rjd3toolkit::autocorrelations.partial(ds, T, n=36)

all<-cbind(ac, pac)
barplot(t(all), beside = T, col = c("red", "blue"), names.arg = c(1:36), ylim=c(-1,1))

spec<-spec.pgram(ds)
```



Arima models

- Auto-projective Models:
  - $x_t = f(x_{t-1}, x_{t-2}, \cdots, \varepsilon_t)$
- ARIMA models (Box-Jenkins)
  - Good approximation of f
  - Flexible, parsimonious

Arima models (definitions)

- Auto-regressive model
  - $\left(1+\varphi_1 B+\cdots+\varphi_p B^p\right) x_t=\Phi(B) x_t=\varepsilon_t$
  - $x_t=\varepsilon_t-\varphi_1 x_{t-1}-\cdots-\varphi_p x_{t-p}$ (≠R, X13)
- Moving average model
  - $x_t=\Theta(B) x_t=\left(1+\theta_1 B+\cdots+\theta_q B^q\right) \varepsilon_t$
  - $x_t=\varepsilon_t+\theta_1 \varepsilon_{t-1}+\cdots+\theta_q \varepsilon_{t-q}$

Arima models (definitions)

- ARMA (p, q)
  - $\Phi(B)x_t = \Theta(B)\varepsilon_t$
- SARIMA$(p, d, q)(bp, bd, bq)_s$
  - $(1 - B)^d(1 - B^s)^{bd}\Phi(B)\Phi_s(B^s)x_t = \Theta(B)\Theta_s(B^s)\varepsilon_t$

- Airline model
  - $(1 - B)(1 - B^s)x_t = (1 + \theta B)(1 + \theta_s B^s)\ \varepsilon_t$

  - Interpretation of $x_t = x_{t-12} + \theta_s\varepsilon_{t-12} + \varepsilon_t$: *same as previous year, partially corrected for the "committed error"*

Arima model properties

- Wold representation of an ARMA model

$$\Phi(B)y_t = \Theta(B)\varepsilon_t \Leftrightarrow y_t = \Psi(B)\varepsilon_t, \Psi(B) = \frac{\Theta(B)}{\Phi(B)}$$

- Auto-covariance generating function (acgf)

$$acgf(y_t) = \sigma^2\Psi(z)\Psi(z^{-1}) = \sigma^2 \frac{\Theta(z)\Theta(z^{-1})}{\Phi(z)\Phi(z^{-1})}$$

- Spectrum ≡ Fourier transform of the acgf

$$g_y(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\Theta\left(e^{-i\lambda}\right)|^2}{|\Phi(e^{-i\lambda})|^2}$$

  - Extension to non-stationary models →Pseudo-spectrum
  - Counterpart of the periodogram, contribution of each frequency to the variance of the series

- Wold representation of an ARMA model

$$\Phi(B)y_t = \Theta(B)\varepsilon_t \Leftrightarrow y_t = \Psi(B)\varepsilon_t, \Psi(B) = \frac{\Theta(B)}{\Phi(B)}$$

- Auto-covariance generating function (acgf)

$$acgf(y_t) = \sigma^2\Psi(z)\Psi(z^{-1}) = \sigma^2\frac{\Theta(z)\Theta(z^{-1})}{\Phi(z)\Phi(z^{-1})}$$

- Spectrum ≡ Fourier transform of the acgf

$$g_y(\lambda) = \frac{\sigma^2}{2\pi}\frac{|\Theta\left(e^{-i\lambda}\right)|^2}{|\Phi(e^{-i\lambda})|^2}$$

Final remarks

- Importance of the representation in the frequency domain for series decomposition
  - Trend ≈ low frequencies
  - Seasonal ≈ seasonal frequencies (related to the periodicity)
  - Irregular ≈ high-frequencies

- Especially true for SEATS