Time Series and Forecasting
ARIMA and SARIMA Models

Maria Eduarda Silva

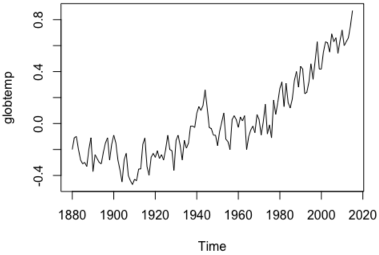
FEP.UP

Recall that

- Time series (a variable observed over time) are correlated data- that shows in the acf
- Stationarity means constant mean, constant variance and autocorrelation that depends only on the lag (how far apart are the observations)
- Stationarity is important to be able to estimate meaningfully (in a statistical sense) the dependence measures: mean, variance and acf
- ARMA models are models for stationary time series that are able to represent (different patterns of) serial correlation in the data

Problem: Nonstationary time series

- However, most of the observed time series are not stationary.
- How do you know that your time series data is not stationary?
  - the first and foremost indication is a mean that is not constant- sometimes it is not easy to check visually
  - the acf decays to zero slowly - observations far apart are highly correlated- this is a consequence of fact that we are not making a statistically meangfull estimate



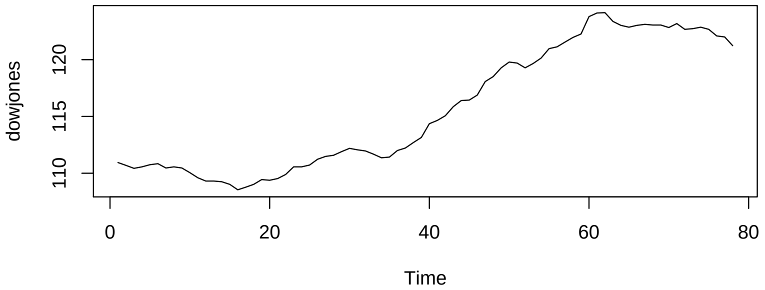
Models for non-stationary data

- In many situations a time series may be thought of as composed by two components $X_t = \mu_t + Y_t$:
  - a time dependent (non-stationary) mean - designated by trend- μ_t
  - a zero mean stationary component Y_t
- Models for μ_t
  - $\mu_t = \beta_0 + \beta_1 t$ - deterministic linear trend
  - $\mu_t = \mu_{t-1} + e_t$ a random walk- slowly varying stochastic trend
  - $\mu_t = \delta + \mu_{t-1} + e_t$ a random wlak with drift- the constant δ induces a linear trend

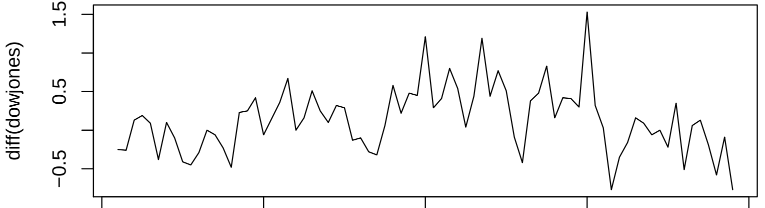
Integrated models for non-stationary data

- In all the above representations $X_t - X_{t-1} = (1-B)X_t = \nabla X_t$ is stationary.
- ∇X_t represents the changes in the varibles on consecutive time unit: dayly , weekly, monthly, quarterly, annual changes (increments)
- Some time but very seldom we need to difference the series d times

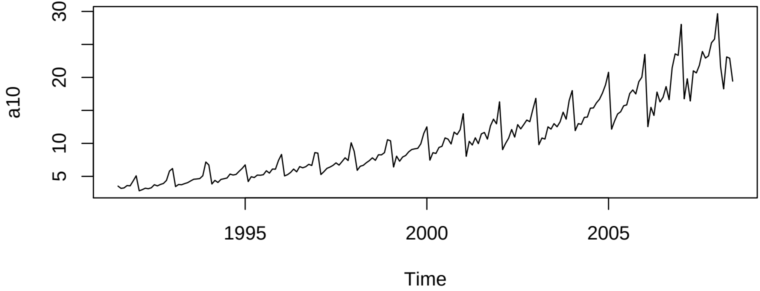
$$X_t = (1-B)^d X_t = \nabla^d X_t$$

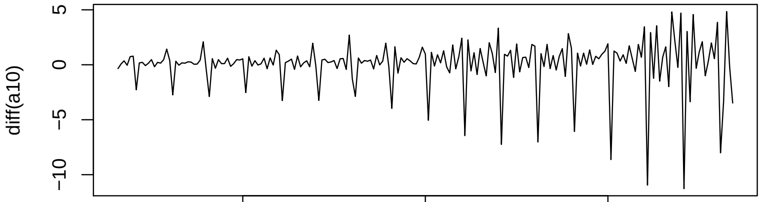


dowjones
110
115
120
0
20
40
60
80
Time



diff(dowjones)
−0.5
0.5
1.5


30
20
10
5
a10
1995
2000
2005
Time


5
0
−5
−10
diff(a10)

ARIMA(p, d, q)

Autoregressive Integrated Moving Average

- Combine ARMA models with differencing
- The reasoning is: the differenced (d times) variables are stationary $Y_t = (1 - B)^d X_t$
- So $(1 - B)^d X_t$ follows a (stationary and invertible) ARMA model:

$$\phi(B) Y_t = \theta(B) e_t$$
$$\phi(B)(1 - B)^d X_t = \theta(B) e_t$$

- X_t is said an ARIMA(p, d, q)

$$\begin{aligned} \text{AR: } p &= \text{order of the autoregressive part} \\ \text{I: } d &= \text{degree of first differencing} \\ \text{MA : } q &= \text{order of the moving average part} \end{aligned}$$

ARIMA(p, d, q)

- White noise: ARIMA(0,0,0)
- Random walk: ARIMA(0,1,0) with no constant
- AR(p): ARIMA(p,0,0)
- MA(q): ARIMA(0,0,q)

ARIMA(p, d, q)

- ARIMA(1,1,1)

$$(1 - a\ B)(1 - B)Y_t = (1 - b\ B)e_t$$

- ARIMA$(1, d, 1)$

$$(1 - a\ B)(1 - B)^d Y_t = (1 - b\ B)e_t$$

$$Y_t = Y_{t-1} + aY_{t-1} - aY_{t-2} + be_{t-1} + e_t$$

- Note that ARIMA models have unit roots on the AR component
- ARIMA(p, d, q)

$$\phi(B)\nabla^d Y_t = \theta(B)e_t$$

$\phi(B)$ and $\theta(B)$ are the AR and MA polynomials

Seasonal Models, SARMA(P,Q)$_S$

SARMA$(P, Q)_S$ where S is the seasonality , ex: S=4 months, S=12 months

$$Y_t = \alpha_1 Y_{t-S} + \ldots + \alpha_P Y_{t-PS} + \beta_1 e_{t-S} + \ldots + \beta_Q e_{t-QS} + e_t$$

- Seasonal Autoregressive polynomial

$$\Phi(z^S) = 1 - \alpha_1 z^S - \ldots - \alpha_P z^{PS}$$

  with roots $z_1, \ldots, z_P$, $|z_i| > 1$ $i = 1, \ldots, P$ for stationarity

- Seasonal MA polynomial

$$\Theta(z^S) = 1 + \beta_1 z^S + \ldots + \beta_Q z^{QS}$$

  with roots $z_1, \ldots, z_P$, $|z_i| > 1$ $i = 1, \ldots, P$

Multiplicative SARMA$(p, q) \times (P, Q)_S$ Models

SARMA$(p, q) \times (P, Q)_S$

$$\phi(B)\Phi(B^S)Y_t = \theta(B)\Theta(B^S)e_t$$

- AR polynomial $\phi(z) = 1 - a_1 z - \ldots - a_p z^p$ with roots $z_1, \ldots, z_p$, $|z_i| > 1 \ i = 1, \ldots, p$
- MA polynomial $\theta(z) = 1 + b_1 z + \ldots + b_q z^q$ with roots $z_1, \ldots, z_q$, $|z_i| > 1 \ i = 1, \ldots, q$
- Seasonal AR polynomial $\Phi(z^S) = 1 - \alpha_1 z^S - \ldots - \alpha_P z^{PS}$ with roots $z_1, \ldots, z_P$, $|z_i| > 1 \ i = 1, \ldots, P$
- Seasonal MA polynomial $\Theta(z) = 1 + \beta_1 z^S + \ldots + \beta_Q z^{QS}$ with roots $z_1, \ldots, z_Q$, $|z_i| > 1 \ i = 1, \ldots, Q$

Examples

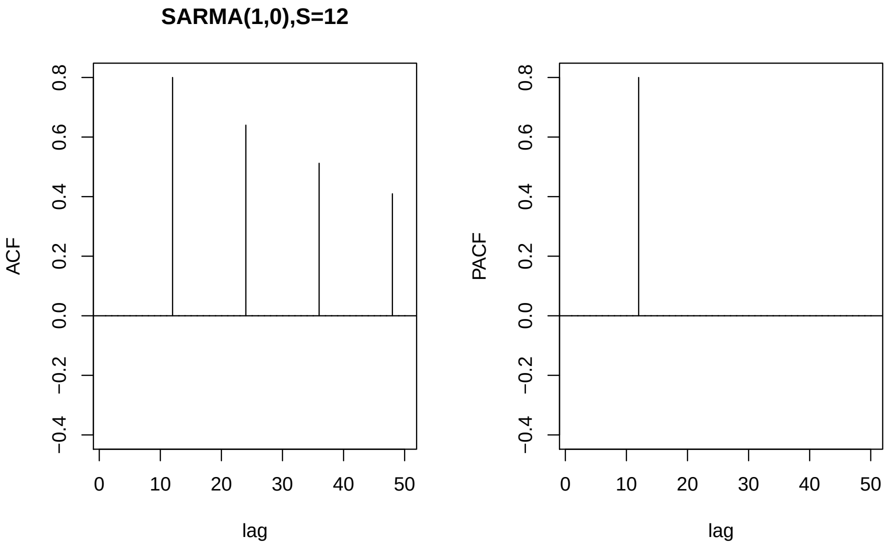
- ARMA(1,1) Seasonal Model, SARMA$(1,1)_{12}$ for monthly data with a yearly component
  $Y_t = \alpha Y_{t-12} + \beta e_{t-12} e_t$ or $(1-\alpha B^{12})Y_t = (1+\beta B^{12})e_t$
- SARMA$(1,1) \times (1,1)_4$ for quarterly data
  $(1-aB)(1-\alpha B^4)Y_t = (1+bB)(1+\beta B^4)e_t$

$|a| < 1$, $|\alpha| < 1$, $|b| < 1$, $|\beta| < 1$

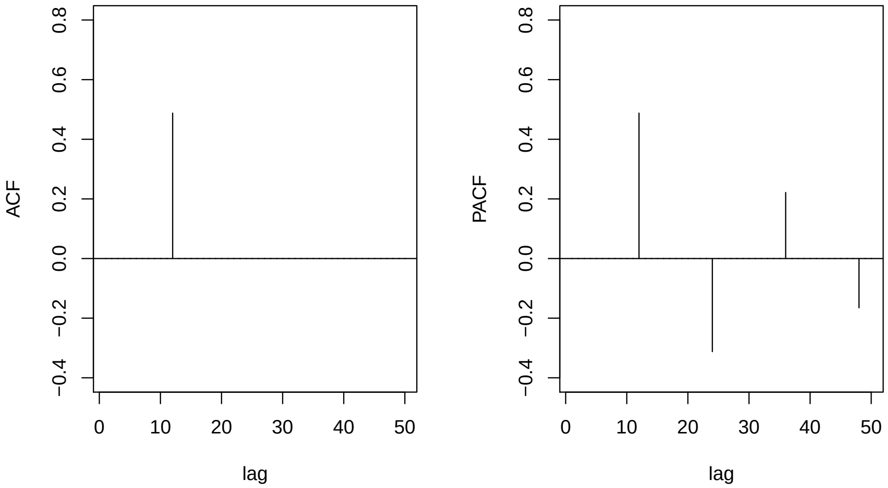
Try the following code for the ACF and PACF of the SARMA

$Y_t = .8Y_{t-12} + e_t + 0.5e_{-1}$

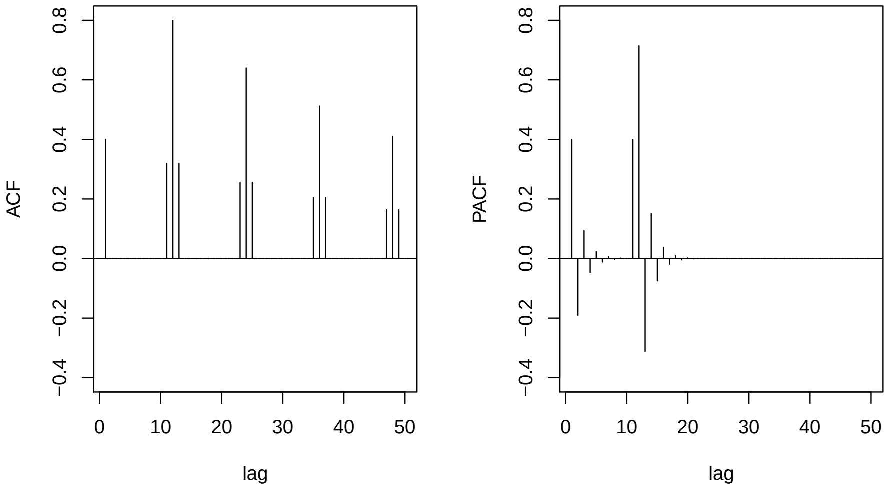
```
phi = c(rep(0,11),.8)
ACF = ARMAacf(ar=phi, ma=0.5, 50)[-1] # [-1] removes 0 lag
PACF = ARMAacf(ar=phi, ma=0.5, 50, pacf=TRUE)
par(mfrow=c(1,2))
plot(ACF, type="h", xlab="lag", ylim=c(-.4,.8)); abline(h=0)
plot(PACF, type="h", xlab="lag", ylim=c(-.4,.8)); abline(h=0)
```


SARMA(1,0),S=12
ACF
PACF
lag
lag
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
10
20
30
40
50

SARMA(0,1),S=12



SARMA(0,1)x(1,0),S=12



Seasonal differences

Many time series present seasonal persistence which occurs when the process is nearly constant in the season. This means that the seasonal component is nearly constant but not quite, that is $S_t \approx S_{t-S}$ instead of $S_t = S_{t-S}$.
We can consider a random walk for the seasonal component $S_t = S_{t-S} + w_t$, meaning a seasonal unit root.
Similarly to ARIMA we define the seasonal differences

$$\nabla_S^D Y_t = (1 - B^S)^D Y_t$$

Usually $D = 1$ is enough
Exemple: $\nabla_{12} Y_t = Y_t - Y_{t-12}$

Multiplicative Seasonal models, SARIMA$(p, d, q)x(P, D, Q)_S$

$$\text{ARIMA} \quad \underbrace{(p, d, q)}_{\uparrow} \quad \underbrace{(P, D, Q)_m}_{\uparrow}$$

$$\begin{pmatrix} \text{Non-} \\ \text{seasonal} \\ \text{part of the} \\ \text{model} \end{pmatrix} \qquad \qquad \begin{pmatrix} \text{Seasonal} \\ \text{part of} \\ \text{the} \\ \text{model} \end{pmatrix}$$

SARIMA example: ARIMA$(0,1,1)x(0,1,1)_4$

$$\text{ARIMA}\quad \underbrace{(p,d,q)}_{\uparrow} \quad \underbrace{(P,D,Q)_m}_{\uparrow}$$

$$\left(\begin{array}{l}\text{Non-}\\ \text{seasonal}\\ \text{part of the}\\ \text{model}\end{array}\right) \qquad \left(\begin{array}{l}\text{Seasonal}\\ \text{part of}\\ \text{the}\\ \text{model}\end{array}\right)$$

where m = number of periods per season.

SARIMA example: ARIMA$(0,1,1)x(0,1,1)_{12}$

$$(1-B)(1-B^{12})Y_t = (1+\theta\ B)(1+\Theta\ B^{12})e_t$$
$$(1-B-B^{12}+B^{13})Y_t = (1+\theta\ B+\Theta\ B^{12}+\theta\Theta\ B^{13})e_t$$
$$Y_t = Y_{t-1}+Y_{t-12}-Y_{t-13}+e_t+\theta e_{t-1}+\Theta e_{t-12}+\theta\Theta e_{t-13}$$

This model provides a reasonable representation for seasonal, nonstationary, economic time series.