Time Series and Forecasting
ARIMA Models

Maria Eduarda Silva
mesilva@fep.up.pt

School of Economics, University of Porto

LMU, July 1 2022

Time Series Models

- Models for the conditional mean
- Models for the conditional variance
- ARMA models: linear models for the conditional mean
- Non-linear models: example Garch model which is a model for conditional variance (won Engle the 2003 Nobel Prize in Economics)

The Autoregressive Model, AR(1)

- If there is dependence in the data we would like a model that allows to predict future outcomes from past outcomes.
- If there is information about Y_t contained on $(Y_{t-1}, Y_{t-2}, \ldots)$ the obvious is to try a regression of Y_t on its lags.
- First we consider a simple model called AR(1).

AR(1)

$$Y_t = c + a\, Y_{t-1} + e_t \tag{1}$$

It is like a simple linear repression with lagged value of Y_t as predictor

The model has two parts:

- $a\ Y_{t-1} \longrightarrow$ that depends on the past
- $e_t \longrightarrow$ the part that is not predictable from the past.
- e_t is iid $N(0, \sigma^2)$
- e_t independent of $Y_{t-1}, Y_{t-2}, \dots$

but we must put restrictions on parameter a

$$|a| < 1$$

so that Y_t is stationary

AR(1)- Some properties

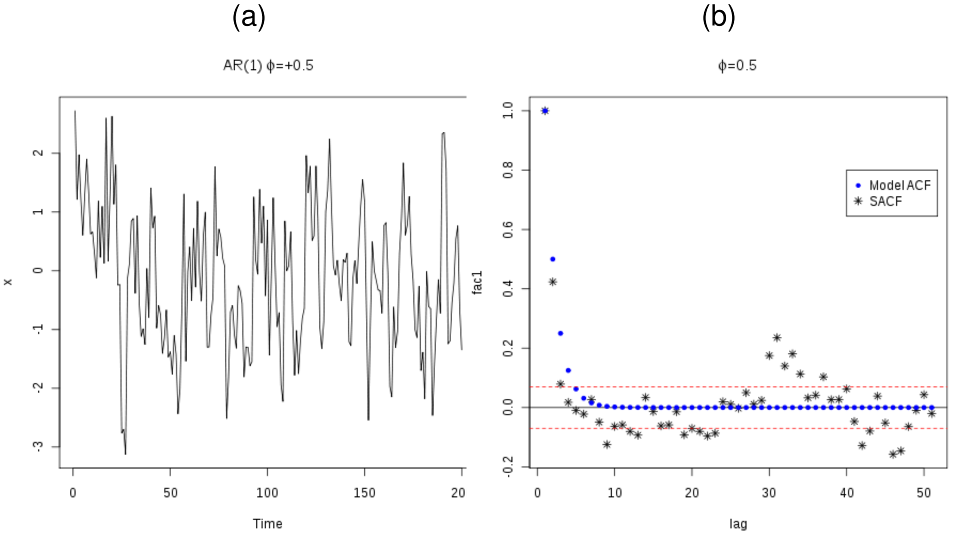
$$Y_t = c + a\ Y_{t-1} + e_t$$

- $E(Y_t) = c/(1-a)$
- $Var(Y_t) = \gamma(0) = \sigma_e^2/(1-a^2)$
- $\gamma(\tau) = a\gamma(\tau-1),\ \ \tau = 2, \ldots$
- $\rho_\tau = a\rho_{\tau-1} = \frac{\gamma(\tau)}{\gamma(0)} = a^{|\tau|}$

Note that a is the correlation between Y_t and Y_{t-1}- successive observations

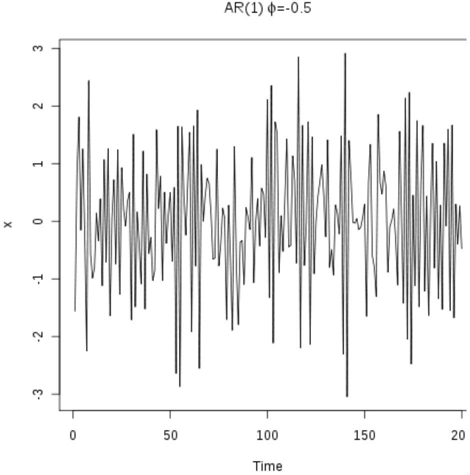
Note that the correlation between Y_t and Y_{t-k} decays exponentially with k

AR(1) ACF and SACF

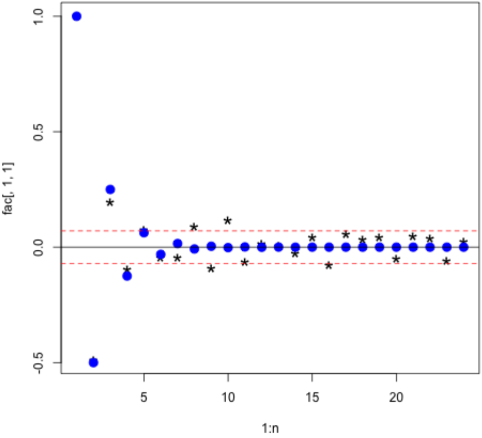


AR(1) ACF and SACF

(a)



(b)



AR(1) and causality

$$
\begin{aligned}
Y_t &= aY_{t-1} + e_t \\
Y_t &= a(aY_{t-2} + e_{t-1}) + e_t = a^2\ Y_{t-2} + a\ e_{t-1} + e_t \\
\ldots & \\
Y_t &= e_t + ae_{t-1} + a^2 e_{t-2} + \ldots \\
Y_t &= \sum_{j=0}^{\infty} a^j e_{t-j}
\end{aligned}
$$

This is called the causal representation of an AR(1) refering to the fact that Y_t does not depend on the future. Check by that $Y_t = \sum_{j=0}^{\infty} a^j e_{t-j}$ satisfies the AR(1) equation.

Compute the mean and autocovariance function of Y_t using the causal representation.

Partial correlation

- X, Y , Z random variables
- Partial correlation between X and Y given Z is obtained by
  - regressing X on Z to obtain the predictor $\hat{X}$, regressing Y on Z to obtain the predictor $\hat{Y}$
  - compute $\rho_{XY|Z} = corr(X - \hat{X}, Y - \hat{Y})$
- Computes the correlation between X and Y with the linear effect of Z removed

The partial autocorrelation function, PACF

- Helps to determine the order of the AR model
- Computes the correlation between Y_t and Y_{t-k} removing the linear effect of the variables between
- The first PACF is the coefficient β_1^1 in the regression $Y_t = \beta_0^1 + \beta_1^1 \; Y_{t-1}$

$$\phi_{11} = \rho(1)$$

- The kth PACF is

$$\phi_{kk} = corr(Y_k - \hat{Y}_k, Y_0 - \hat{Y}_0)$$

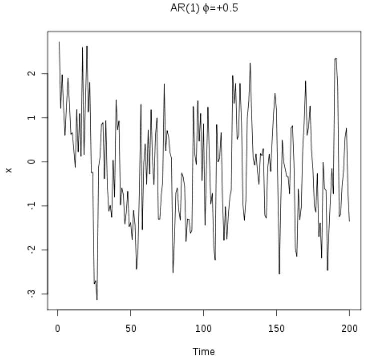
where $\hat{Y}_t$ is the predictor of Y_t based on $\{Y_1, \ldots, Y_{k-1}\}$, $k = 2, 3, \ldots$

PACF of AR(1) and sample PACF, SPACF

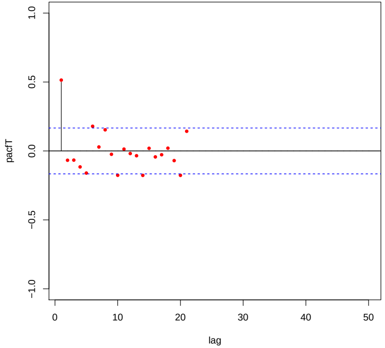
- For an AR(1)
  - $\phi_{11} = a$
  - $\phi_{hh} = 0, \quad h \geq$
- A sample PACF β_k^k is significant if $\beta_k^k > 2/\sqrt{n}$ (Bartlett test)

AR(1) PACF and SPACF

(a)

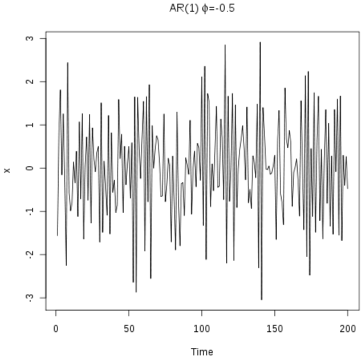


(b)

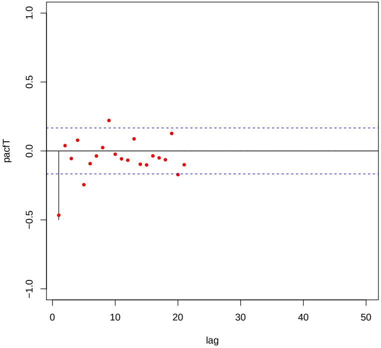


AR(1) PACF and SPACF

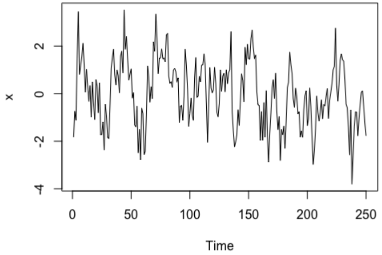
(a)

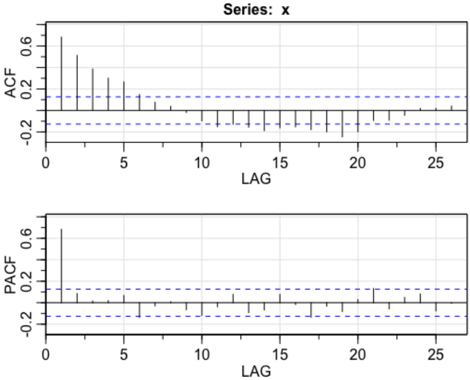


(b)



AR(1) $a = 0.7$ PACF and SPACF





AR(p)

- Generalization of AR(1) model: the dependence is on p past values

$$Y_t = c + a_1 Y_{t-1} + \ldots + a_p Y_{t-p} + e_t$$

  Multiple regression with p lagged values of Y_t as predictors
- e_t iid sequence with variance σ_e^2
- Conditions on the a_i so that the model exists (in the mathematical sense) - the (complex) roots $z_1, \ldots, z_p$ of the autoregressive polinomial

$$\phi(z) = 1 - a_1 z - \ldots - a_p z^p$$

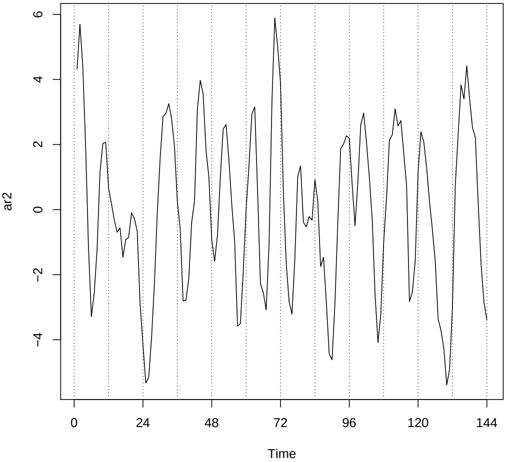
  are $|z_i| > 1$
- allows periodic behaviour

Example: AR(2)

$$Y_t = 1.5Y_{t-1} - 0.75Y_{t-2} + e_t$$

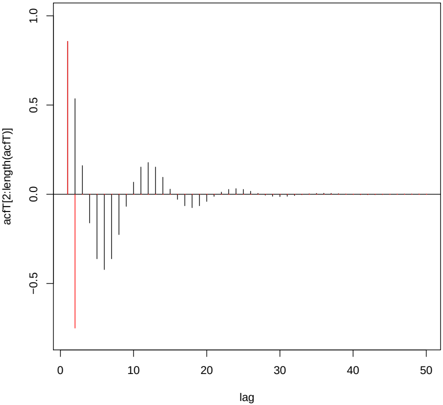
```
z=c(1,-1.5,0.75)
polyroot(z)
[1] 1+0.57735i 1-0.57735i
abs(polyroot(z))^2
[1] 1.333333 1.333333
```

Realization of $Y_t = 1.5Y_{t-1} - 0.75Y_{t-2} + e_t$

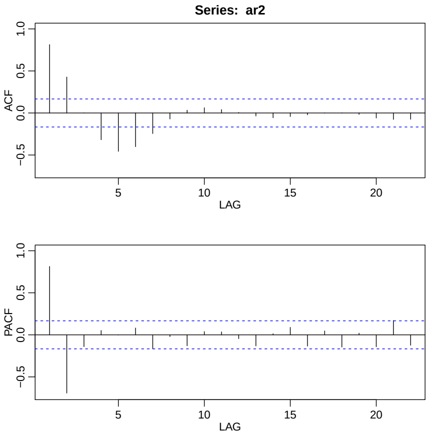


AR(2) ACF, PACFand sample counterparts

(a)



(b)



MA(1) Model

- $$X_t = e_t + be_{t-1} = (1 - b\ B)e_t$$

  where $|b| < 1$ and e_t iid var(e_t)=σ_e^2
- Stationary for all b
- The model is different from AR(1) since we write X_t as a sum of two random draws from a distribution (usually normal)
- The value of X today depends on the *surprise* from yesterday and the *surprise* from today
- ACF: $\rho(1) = b/(1 + b^2)$ e $\rho(k) = 0,\ k \geq 2$

MA(q) model

- $$X_t = e_t + b_1 e_{t-1} + \ldots + b_q e_{t-q} \quad (2)$$

  where e_t iid random variables N(0,1) and the b_j are such that the roots $z_1, \ldots, z_q$ of the polinomial

  $$\theta(z) = 1 + b_1 z + \ldots + b_q z^q$$

  satisfy $|z_i| > 1$, $i = 1, \ldots, q$ (for technical reasons)
- The ACF is zero for lags $\tau > q$ and the PACF decays to zero.
- the joint use of ACF and PACF allows to distinguish between AR and MA models

ACF e PACF properties for AR(p) and MA(q) models

	AR(p)	**MA(q)**
ACF	Decay to zero	Zero for lags $\geq q+1$
PACF	Zero for lags $\geq p+1$	Decay to Zero

ARMA models

- Data is generally not well represented by AR or MA models
- Combine both $\rightarrow$ ARMA models
- ARMA(1,1)

$$X_t = c + aX_{t-1} + be_{t-1} + e_t$$

e_t iid, $|a| < 1$, $|b| < 1$

ARMA models

- ARMA(p, q)

$$X_t = c + a_1 X_{t-1} + \ldots + a_p X_{t-p} + b_1 e_{t-1} + \ldots + b_q e_{t-q} + e_t$$

where $a_1, \ldots, a_p$ and $b_1, \ldots, b_q$ are such that the roots of the polynomials

$$\phi(z) = 1 - a_1 z - \ldots - a_p z^p$$

and

$$\theta(z) = 1 + b_1 z + \ldots + b_q z^q$$

are inside the unit circle, $|z_i| > 1$

- Predictors include both **lagged values of X_t lagged vaues of errors/innovations.**

AutorRegressive Integrated Moving Average models

- Combine ARMA models with differencing
- d-differenced series follows an ARMA model
  - Difference the series d times

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

  - ARMA(p, q) model for Y_t

$$X_t = a_1 X_{t-1} + \ldots + a_p X_{t-p} + b_1 e_{t-1} + \ldots + b_q e_{t-q} + e_t$$

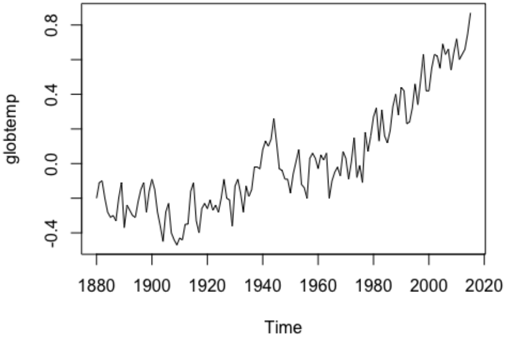
- Need to choose p, q, d and whether include constant

ARIMA(p, d, q)

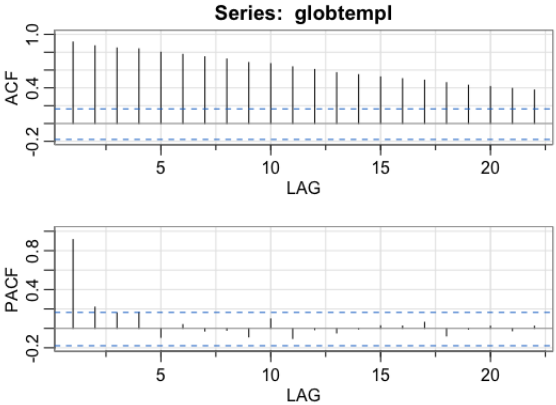
- White noise: ARIMA(0,0,0)
- Random walk: ARIMA(0,1,0) with no constant
- Random walk with drift: ARIMA(0,1,0) with a constant
- AR(p): ARIMA$(p, 0, 0)$
- MA(q): ARIMA$(0, 0, q)$
- ARMA(p, q): ARIMA$(p, 0, q)$
- ARIMA(1,1,1)

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

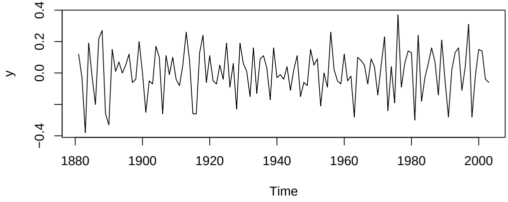
Global Temperature

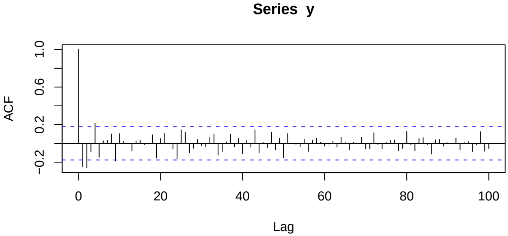


Global Temperature

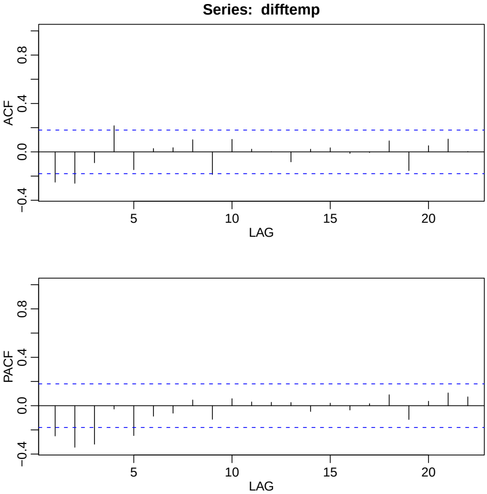


Differenced global Temperature





ACF and PACF of Differenced global Temperature



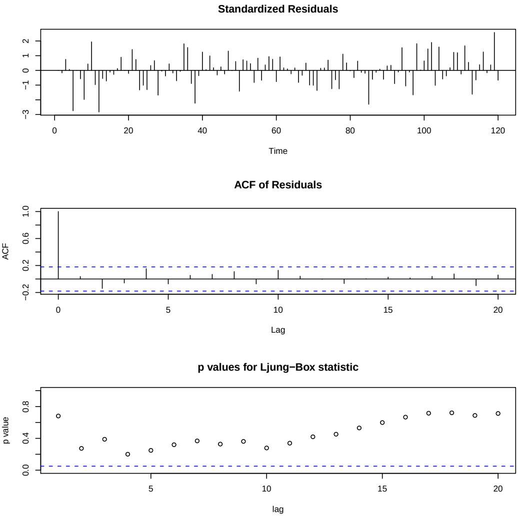
Global Temperature - ARIMA(1,1,1)?

Coefficients:

	ar1	ma1	constant
	0.2707	-0.8367	0.0053
s.e.	0.1144	0.0633	0.0028

what about the residuals?

Residuals from ARIMA(1,1,1) for Temperature data



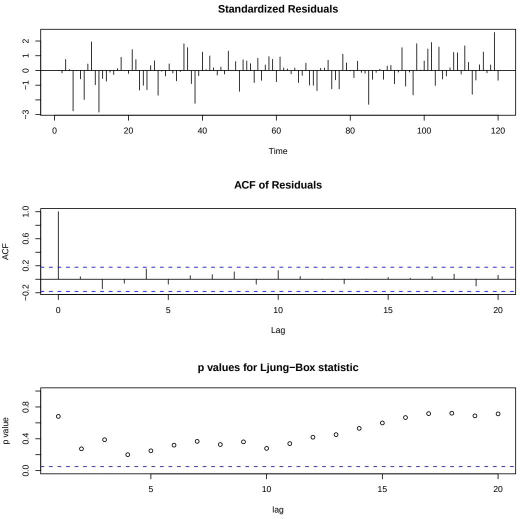
What about an ARIMA(1,1,2)?

Coefficients:

	ar1	ma1	ma2	constant
	-0.0415	-0.4641	-0.2713	0.0061
s.e.	0.2257	0.2068	0.1440	0.0030

what about the residuals?

Residuals from ARIMA(1,1,2) for Temperature data



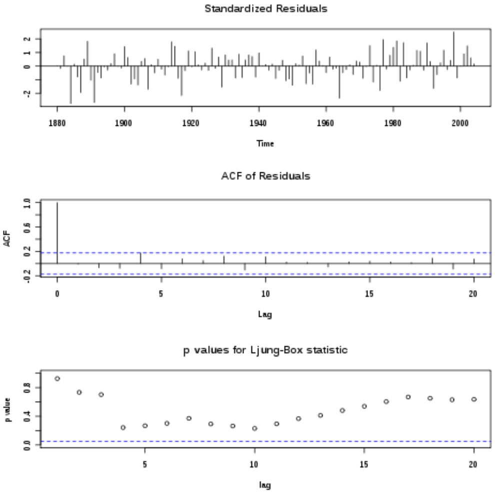
An ARIMA(0,1,2)?

Coefficients:

ma1	ma2	constant
-0.4992	-0.2486	0.0061
s.e. 0.0843	0.0796	0.0030

what about the residuals?

Residuals from ARIMA(0,1,2) for Temperature data



How to choose between ARIMA(1,1,1) and ARIMA(0,1,2)?

Information Criteria

Akaike (1969, 1973, 1974) suggested measuring the goodness of a model by balancing the error of the fit against the number of parameters in the model. Thus Akaike Information Criteria was born. Later developed into AICc and BIC (Bayesian Information Criteria).

Definition 2.1 Akaike's Information Criterion (AIC)

$$\mathrm{AIC} = \log\ \widehat{\sigma}_k^2 + \frac{n+2k}{n}, \tag{2.16}$$

where $\widehat{\sigma}_k^2$ is given by (2.15) and k is the number of parameters in the model.

Information criteria

For comparing (nested) models leading to automatic choice of order:
Good models are obtained by minimizing either

AIC Akaike Information Criteria

$$\mathrm{AIC} = -2\log(L) + 2(p+q+k+1)$$

where L is the likelihood of the data and $k = 1$ if we are also estimating the mean

$\mathrm{AIC_C}$ Corrected AIC

$$\mathrm{AIC}_C = \mathrm{AIC} + \frac{2(p+q+k+1)(p+q+2)}{n-p-q-k-2}$$

BIC Bayesian Information Criteria

$$\mathrm{BIC} = \mathrm{AIC} + \log(n)(p+q+k+1)$$

How to choose between ARIMA(1,1,1) and ARIMA(0,1,2)?

Given more than one adequate model choose the model that minimizes AIC or AICc

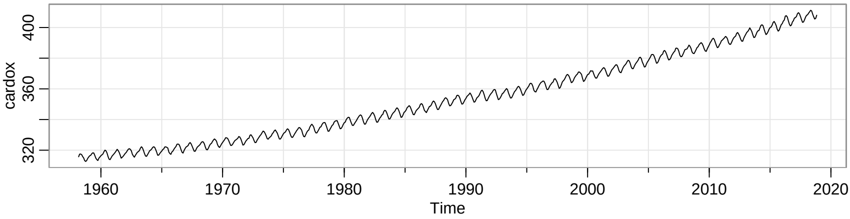
Model	AIC	AICc
ARIMA(1,1,1)	-3.028143	-3.008577
ARIMA(0,1,2)	-3.044159	-3.024593

Seasonal ARIMA Models

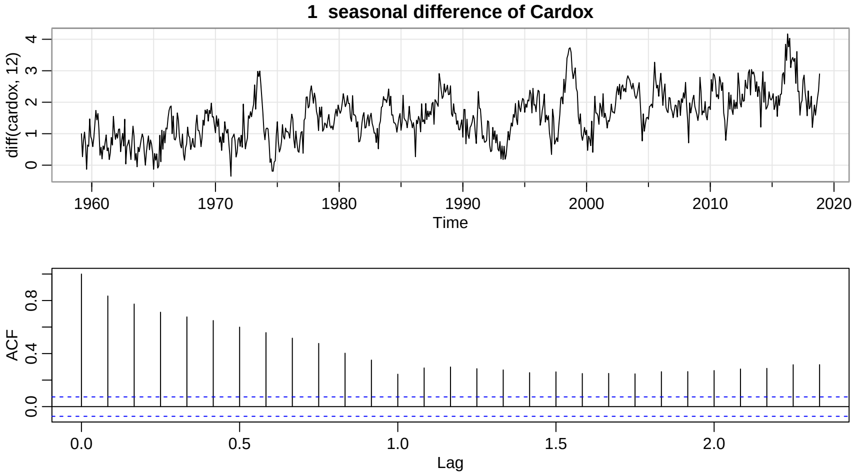
- Seasonal and non-seasonal terms combine multiplicatively
- SARIMA$(p, d, q) \times (P, D, Q)_S$ where
  - d differences; D seasonal differences
  - p AR lags; q MA lags
  - P Seasonal AR lags; Q Seasonal MA lags
  - S is the seasonality , ex: S=4 months, S=12 months
  - This model provides a reasonable representation for seasonal, nonstationary, economic time series.
-

ARIMA	$\underbrace{(p,d,q)}$	$\underbrace{(P,D,Q)_m}$
	$\uparrow$ Non-seasonal part of the model	$\uparrow$ Seasonal part of of the model

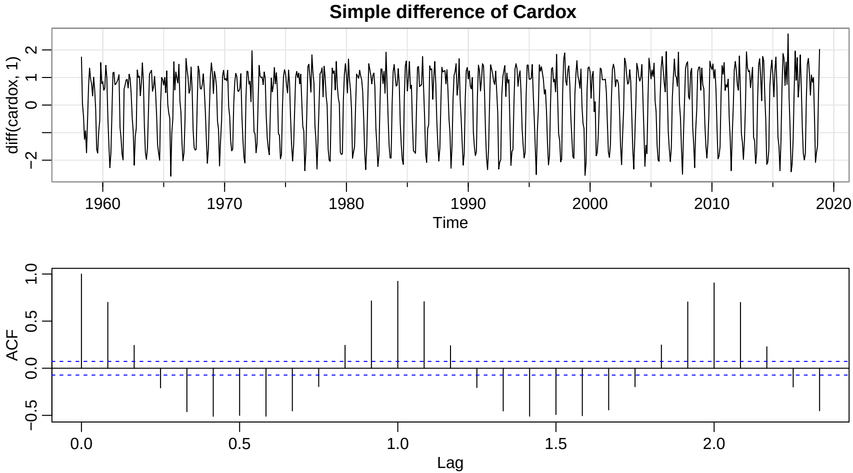
Modelling CARDOX



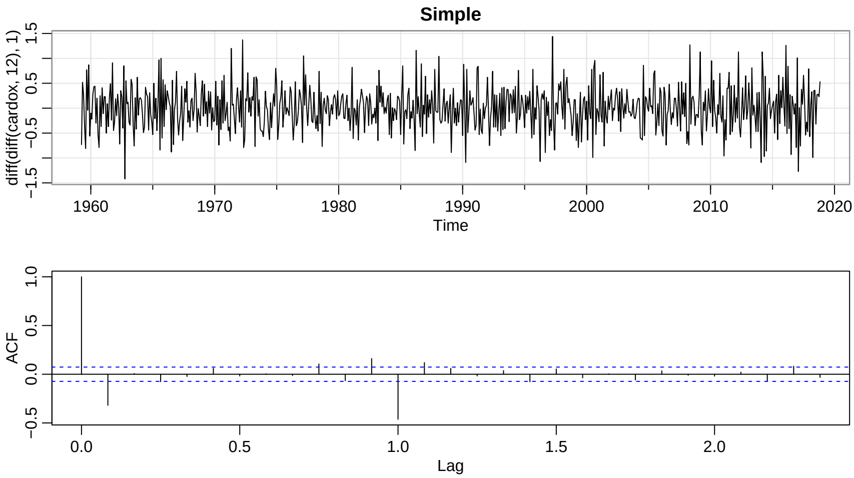
Seasonal Difference of CARDOX



Simple difference of CARDOX



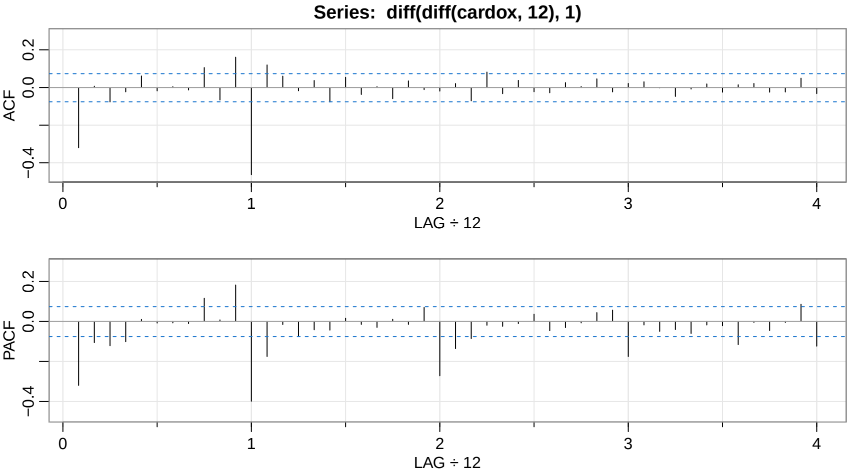
Simple and Seasonal difference of CARDOX



ACF and PACF of differenced CARDOX

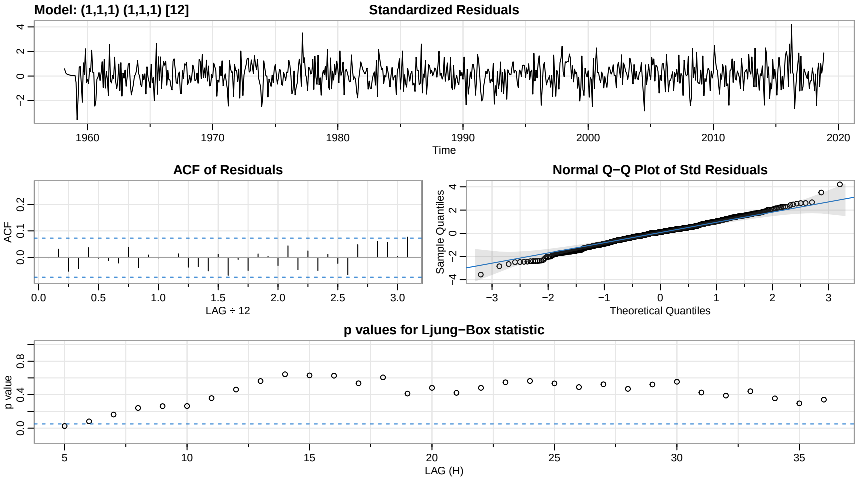
scale=0.5]acf-and-pacf-of-differenced-cardox

```
acf2(diff(diff(cardox,12),1))
```



CARDOX

```
cardox_fit1 = sarima(cardox,1,1,1,1,1,1,12)
```



CARDOX

```
cardox_fit1
```

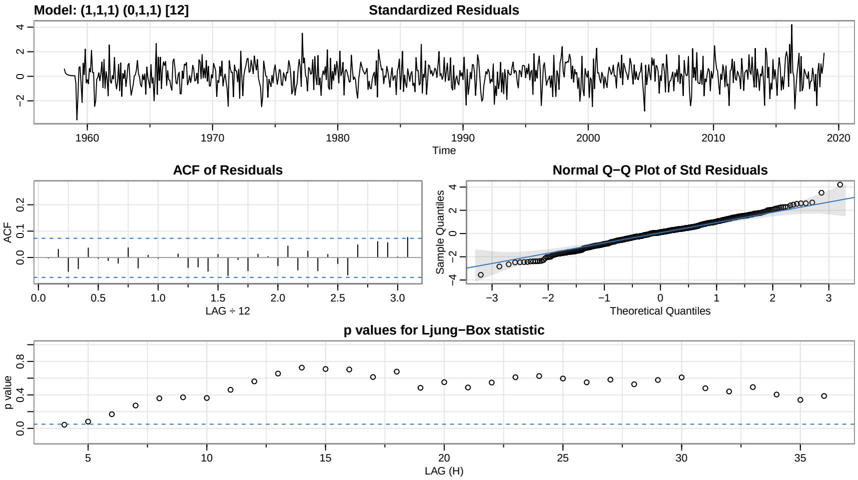
```
## $fit
##
## Coefficients:
##          ar1      ma1      sar1      sma1
##       0.1941  -0.5578  -0.0008  -0.8647
## s.e.  0.0953   0.0813   0.0427   0.0211
##
## sigma^2 estimated as 0.09585:  log likelihood = -184.82,  a
##
## $degrees_of_freedom
## [1] 712
##
```

CARDOX

```
## $ttable
##      Estimate     SE  t.value p.value
## ar1    0.1941 0.0953   2.0364  0.0421
## ma1   -0.5578 0.0813  -6.8645  0.0000
## sar1  -0.0008 0.0427  -0.0185  0.9852
## sma1  -0.8647 0.0211 -40.9625  0.0000
##
## $AIC
## [1] 0.5302324
##
## $AICc
## [1] 0.530311
##
## $BIC
## [1] 0.5621715
```

Cardox

```
cardox_fit2=sarima(cardox,1,1,1,0,1,1,12, no.constant = TRUE)
```



CARDOX

```
cardox_fit2
```

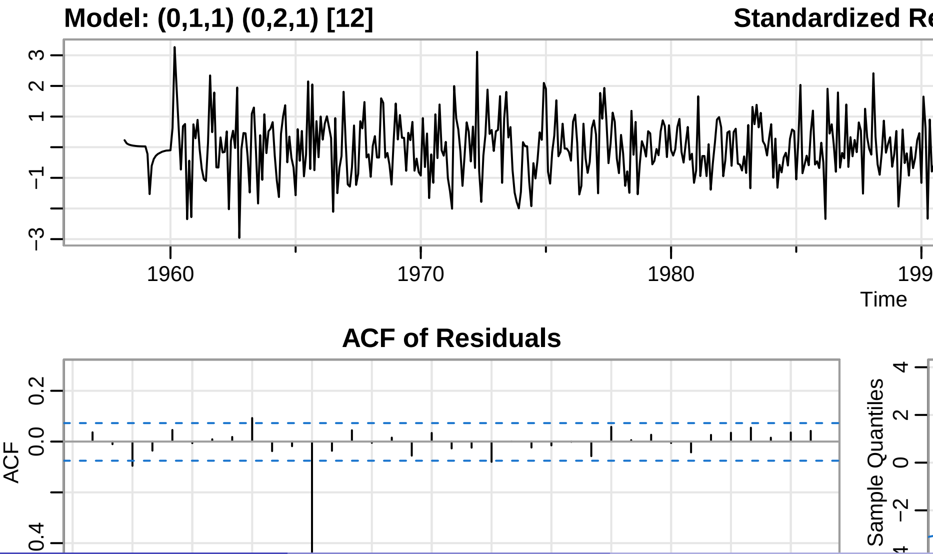
```
## Coefficients:
##          ar1      ma1     sma1
##       0.1941  -0.5578  -0.8648
## s.e.  0.0953   0.0813   0.0189
##
## sigma^2 estimated as 0.09585:  log likelihood = -184.82,  a
##
## $degrees_of_freedom
## [1] 713
```

CARDOX

```
## $ttable
##      Estimate     SE  t.value p.value
## ar1    0.1941 0.0953   2.0374   0.042
## ma1   -0.5578 0.0813  -6.8634   0.000
## sma1  -0.8648 0.0189 -45.7161   0.000
##
## $AIC
## [1] 0.5274396
##
## $AICc
## [1] 0.5274867
##
## $BIC
## [1] 0.5529909
```

Cardox

```
cardox_fit3=sarima(cardox,0,1,1,0,2,1,12, no.constant = TRUE)
```

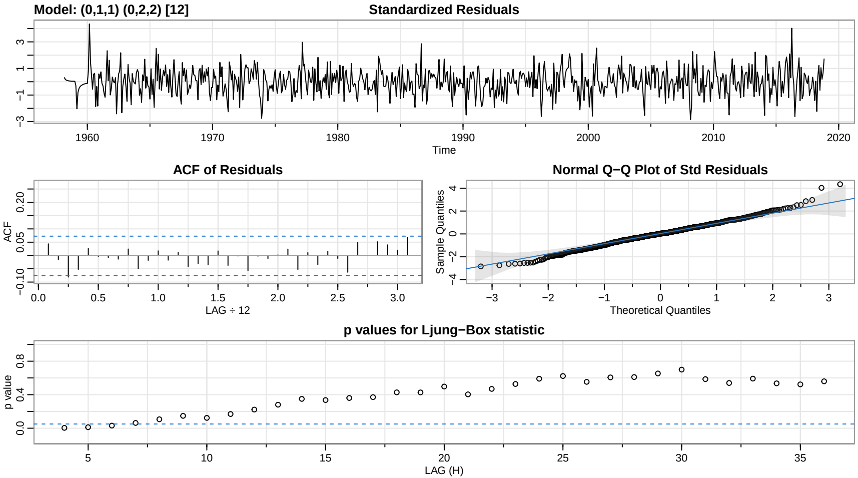


CARDOX

```
## $ttable
##      Estimate     SE  t.value p.value
## ma1   -0.4008 0.0382 -10.4883       0
## sma1  -0.9992 0.0125 -80.0660       0
##
## $AIC
## [1] 1.167823
##
## $AICc
## [1] 1.167847
##
## $BIC
## [1] 1.187241
```

Cardox

```
cardox_fit3=sarima(cardox,0,1,1,0,2,2,12, no.constant = TRUE)
```




```
cardox_fit3
```

Cardox

```
## $fit
## Coefficients:
##          ma1     sma1    sma2
##      -0.4153  -1.9038  0.9095
## s.e.   0.0388   0.0268  0.0264
##
## sigma^2 estimated as 0.09724:  log likelihood = -234.14,  a
##
## $degrees_of_freedom
## [1] 701
##
```

Cardox

```
## $ttable
##       Estimate         SE            t.value          p.value
## ma1    -0.4153   0.0388   -10.6930       0
## sma1  -1.9038  0.0268   -70.9709       0
## sma2   0.9095   0.0264      34.3908       0
##
## $AIC
## [1] 0.6765383
##
## $AICc
## [1] 0.676587
##
## $BIC
## [1] 0.7024291
```

A model for CARDOX

Model	Residuals	Parameters	AICc	BIC
$(1,1,1) \times (1,1,1)$	✓		0.5303	0.5622
$(1,1,1) \times (0,1,1)$	✓	✓	**0.5275**	**0.5530**
$(1,1,1) \times (0,2,1)$	correlated	near unit root	1.167823	1.1872
$(1,1,1) \times (0,2,2)$	✓	near unit root	0.6766	0.7024

Lab Session

Lab Session 2: choose a time series and find a suitable SARIMA model

The file LabSession21.html and LabSession22.html provide examples.

References

- R. Shumway, D. Stoffer. Time series analysis and its applications with R examples. Springer Texts in Statistics, 2006.
- D. Stoffer web page *http://www.stat.pitt.edu/stoffer/dss.html* data, tips about R *http://www.stat.pitt.edu/stoffer/tsa2/*
- R. Hyndman web page *http://robjhyndman.com/* data, slides about R, interesting blog, lots of stuff about forecasting