Time series analysis

Agathe Guilloux

Organizational issues

- 3 classes : 02/06, 02/20 and 03/06
- 2 Labs. The solutions of the practical labs have to be submitted and will be graded.
  - 1st on 02/27 for ENSIIE students and 03/02 for M1 MINT students
  - 2nd on 03/20 for ENSIIE students and 03/23 for M1 MINT students
- Slides, R examples and labs are on my webpage http ://www.math-evry.cnrs.fr/members/aguilloux/enseignements/timeseries
- my email agathe.guilloux@math.cnrs.fr

Chapter 1 : Time series characteristics

Johnson & Johnson quaterly earnings [SS10] I

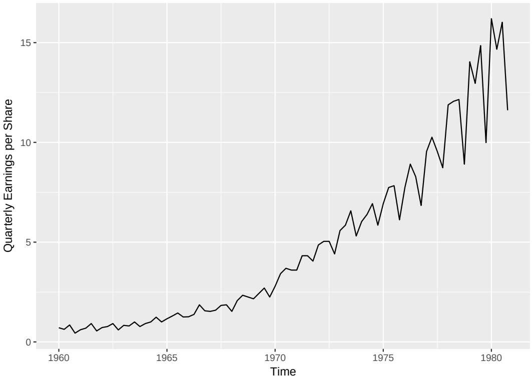


FIG.: Quarterly earnings per share for the U.S. company Johnson & Johnson.

Notice :

- the increasing underlying trend and variability,
- and a somewhat regular oscillation superimposed on the trend that seems to repeat over quarters.

Johnson & Johnson quaterly earnings [SS10] II

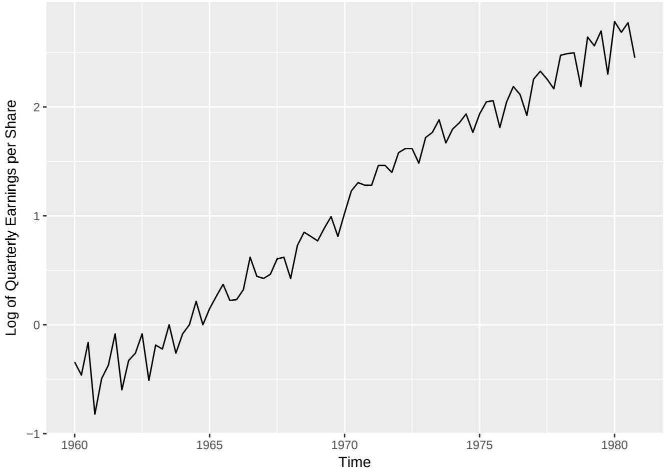


FIG.: Quarterly log(earnings) per share for the U.S. company Johnson & Johnson.

Notice :

- the trend is now (almost) linear

Global temperature index from 1880 to 2015 I

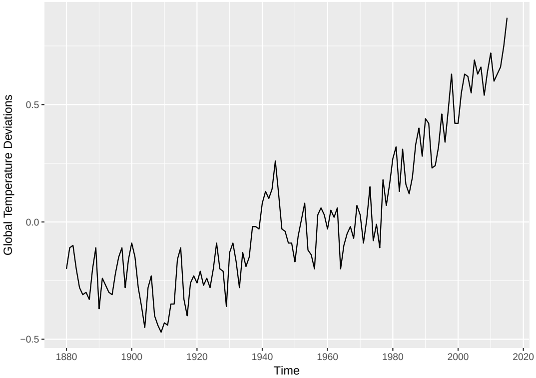


FIG. : Global temperature deviation (in $^\circ C$) from 1880 to 2015, with base period 1951-1980.

Notice :

- the trend is not linear (with periods of leveling off and then sharp upward trends).

Speech data

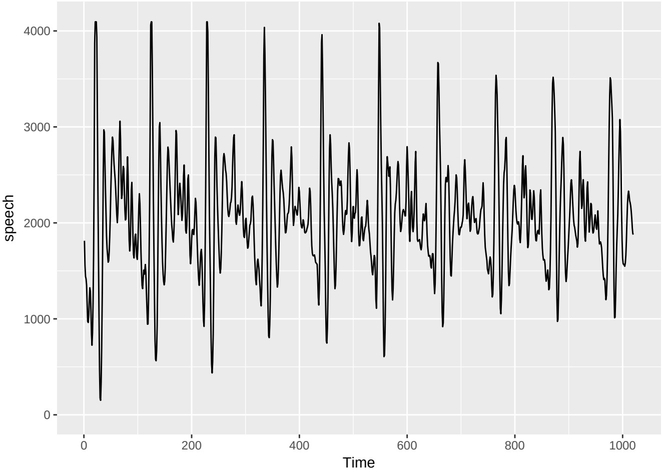


FIG. : Speech recording of the syllable "aaa ... hhh" sampled at 10,000 points per seconds with $n = 1020$ points [SS10]

Notice :

- the repetition of small wavelets.

Dow Jones Industrial Average I

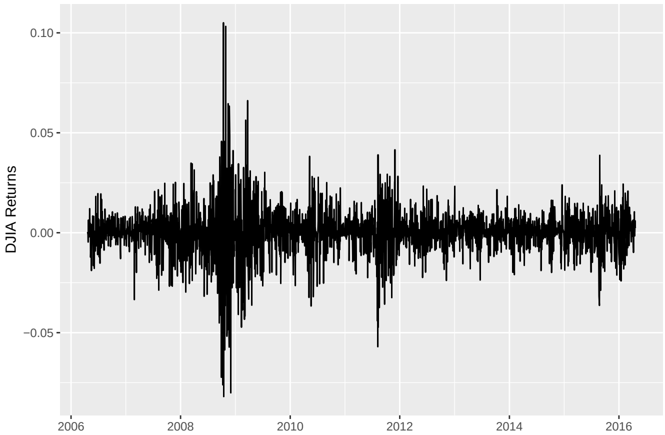


Fig. : Dayly percent change of the Dow Jones Industrial Average from April 20,2006 to April 20,2016 [SS10]

Dow Jones Industrial Average II

Notice :

- the mean of the series appears to be stable with an average return of approximately zero,
- the volatility (or variability) of data exhibits clustering ; that is, highly volatile periods tend to be clustered together.

El Niño and Fish Population I

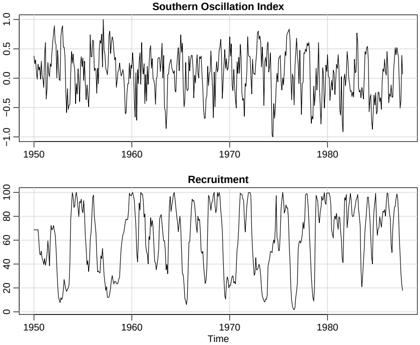


FIG. : Monthly values of an environmental series called the Southern Oscillation Index and associated Recruitment (an index of the number of new fish). [SS10]

Notice :

- SOI measures changes in air pressure related to sea surface temperatures in the central Pacific Ocean.

El Niño and Fish Population II

- The series show two basic oscillations types, an obvious annual cycle (hot in the summer, cold in the winter), and a slower frequency that seems to repeat about every 4 years.
- The two series are also related ; it is easy to imagine the fish population is dependent on the ocean temperature.

fMRI Imaging I

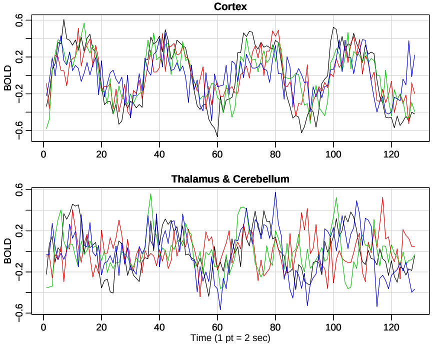


FIG. : Data collected from various locations in the brain via functional magnetic resonance imaging (fMRI) [SS10]

- Notice the periodicities.

What we are seeking

To construct models

- to describe
- to forecast

times series.

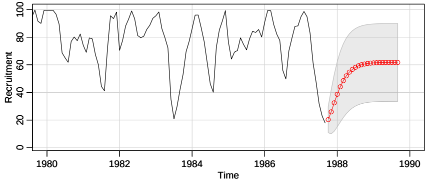


FIG.: Twenty-four month forecats for the Recruitment series shown on slide 11

Times series models

Time series and model

A **time series** is a sequence $(X_t)_{t\in\mathbb{Z}}$ of r.v., i.e. a stochastic process.
A **time series model** specifies (at least partially) the joint distribution of the sequence.

Notation : when no confusion is possible, we'll write for short X instead of $(X_t)_{t\in\mathbb{Z}}$.

White noise

$(\omega_t) \sim WN(0, \sigma^2)$ when

- $\text{Cov}(\omega_s, \omega_t) = 0 \ \forall s, t \in \mathbb{Z}$
- $\mathbb{E}(\omega_t) = 0 \ \forall t \in \mathbb{Z}$
- $\mathbb{V}(\omega_t) = \sigma^2 \ \forall t \in \mathbb{Z}$

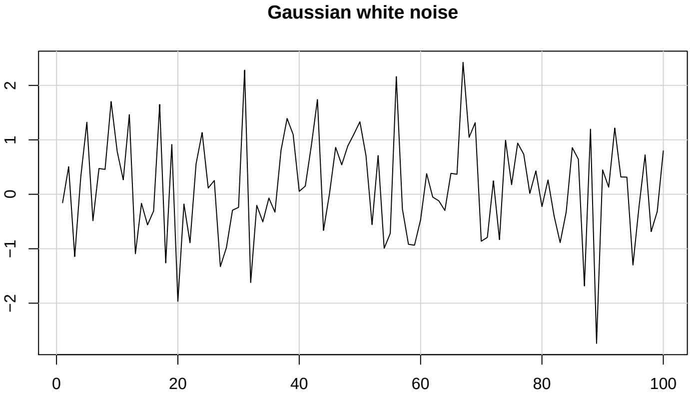
Notation : for white noises, greek letters will be used ω, η

i.i.d. white noise and i.i.d. Gaussian white noise

$(\omega_t) \sim i.i.d.(0, \sigma^2)$ when

- $(\omega_t) \sim WN(0, \sigma^2)$
- and (ω_t) are i.i.d.

$(\omega_t) \sim i.i.d.\mathcal{N}(0, \sigma^2)$ if, in addition, $\omega_t \sim \mathcal{N}(0, \sigma^2)$ for all $t \in \mathbb{Z}$.

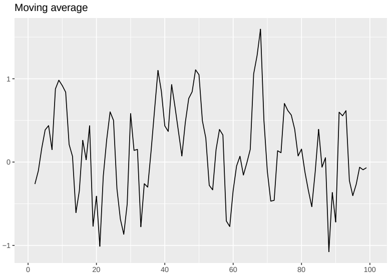


Models with serial correlation I

Moving averages

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}}$ and define the series $(X_t)_{t\in\mathbb{Z}}$ as

$$X_t = \frac{1}{3}(\omega_{t-1} + \omega_t + \omega_{t+1})\ \forall t \in \mathbb{Z}.$$



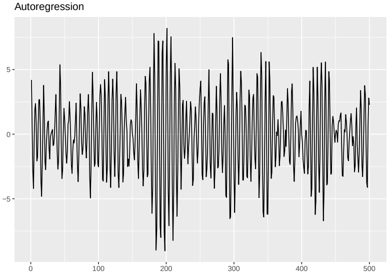
Notice the similarity with SIO and some fMRI series.

Models with serial correlation II

Autoregression

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}}$ and define the series $(X_t)_{t\in\mathbb{Z}}$ as

$$X_t = X_{t-1} - 0.9X_{t-2} + \omega_t \;\forall t \in \mathbb{Z}.$$



Notice

- the almost periodic behavior and the similarity with the speech series example
- the above definition misses initial conditions, we'll come back on that later.

Models with serial correlation III

Random walk with drift

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}}$ and define the series $(X_t)_{t\in\mathbb{Z}}$ as

$$X_t = \underbrace{\delta}_{\text{drift}} + \underbrace{X_{t-1}}_{\text{previous position}} + \underbrace{\omega_t}_{\text{step}} \quad \forall t \in \mathbb{Z},$$

with $X_0 = 0$

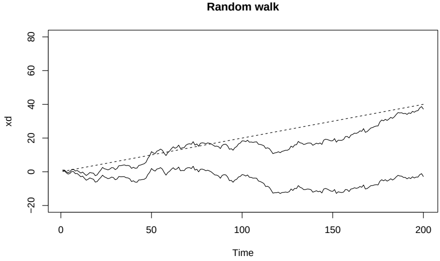


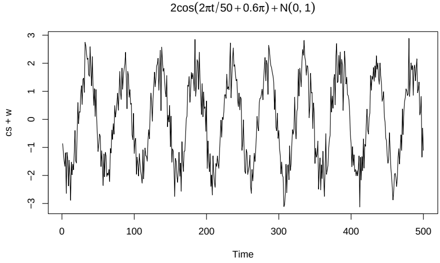
FIG. : Random walk with drift $\delta = 0.2$ (upper jagged line), with $\delta = 0$ (lower jagged line) and line with slope 0.2 (dashed line)

Models with serial correlation IV

Signal plus noise

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}}$ and define the series $(X_t)_{t\in\mathbb{Z}}$ as

$$X_t = \underbrace{2\cos\left(2\pi\frac{t+15}{50}\right)}_{\text{signal}} + \underbrace{\omega_t}_{\text{noise}}$$



Notice the similarity with fMRI signals.

Measures of dependence

We now introduce various measures that describe the **general behavior of a process as it evolves over time.**

Mean measure

Define, for a time series $(X_t)_{t\in\mathbb{Z}}$, the mean function

$$\mu_X(t) = \mathbb{E}(X_t) \; \forall t \in \mathbb{Z}$$

when it exists.

Exercise

Compute the mean functions of

- the moving average defined in slide 17.
- the random walk plus drift defined in slide 19
- the signal+noise model of slide 20

Autocovariance

We now assume that for all $t \in \mathbb{Z}$, $X_t \in \mathbb{L}^2$.

Autocovariance

The autocovariance function of a time series $(X_t)_{t\in\mathbb{Z}}$ is defined as

$$\gamma_X(s,t) = \mathrm{Cov}(X_s, X_t) = \mathbb{E}\big((X_s - \mathbb{E}(X_s))(X_t - \mathbb{E}(X_t))\big)$$

for all $s, t \in \mathbb{Z}$. It is a symmetric function $\gamma_X(s,t) = \gamma_X(t,s)$.

It measures the linear dependence between two values of the same series observed at different times.

Exercise

Compute the autocovariance functions of

- the white noise defined in slide 15
- the moving average defined in slide 17

Autocorrelation function (ACF)

Autocorrelation function

The ACF of a time series $(X_t)_{t\in\mathbb{Z}}$ is defined as

$$\rho_X(s,t) = \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}}$$

for all $s, t \in \mathbb{Z}$. Is is a symmetric function.

Stationarities

Strict stationarity

A time series (X_t) is **strictly stationary** if for all $k \geq 1$, $t_1, \ldots, t_k$ and $h \in \mathbb{Z}$

$$\mathcal{L}(X_{t_1}, \ldots, X_{t_k}) = \mathcal{L}(X_{t_1+h}, \ldots, X_{t_k+h})$$

Weak stationarity

A time series (X_t) is **weakly stationary** if

- μ_X is independent of t and
- $h \mapsto \gamma_X(t+h, t)$ is independent ot t.

In this case, we write $\gamma_X(h)$ as short for $\gamma_X(h, 0)$.

Exercise

Check the stationarity of the following processes :

- the white noise, defined on slide 15
- the random walk, defined on slide 19

Theorem

The autocovariance function γ_X of a stationary time series X verifies

1. $\gamma_X(0) \geq 0$
2. $|\gamma_X(h)| \leq \gamma_X(0)$
3. $\gamma_X(h) = \gamma_X(-h)$
4. γ_X *is positive-definite.*

Furthermore, any function γ that satisfies (3) and (4) is the autocovariance of some stationary time series.

Reminder : A function $f : \mathbb{Z} \mapsto \mathbb{R}$ is positive-definite if for all n, the matrix F_n, with entries $(F_n)_{i,j} = f(i-j)$, is positive definite. A matrix $F_n \in \mathbb{R}^{n \times n}$ is positive-definite if, for all vectors $a \in \mathbb{R}^n$, $a^\top F_n a \geq 0$.

Moving average MA(1) model

Warning : not to be confused with moving average of slide 17.

Moving average model MA(1)

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}} \sim WN(0\,,\sigma^2)$ and construct the MA(1) as

$$X_t = \omega_t + \theta\omega_{t-1} \ \forall t \in \mathbb{Z}$$

Exercise

- Is it stationary ?
- Compute its ACF.

Autoregressive AR(1) model

Autoregressive AR(1)

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}} \sim WN(0\,,\sigma^2)$ and construct the AR(1) as

$$X_t = \phi X_{t-1} + \omega_t \;\forall t \in \mathbb{Z}$$

Exercise

Assume that it is stationary and compute

- its mean function
- its ACF.

Linear processes

Linear process

Consider a white noise $(\omega_t)_{t\in\mathbb{Z}} \sim WN(0\,,\sigma^2)$ and define the linear process X as follows

$$X_t = \mu + \sum_{j\in\mathbb{Z}} \psi_j \omega_{t-j} \;\forall t \in \mathbb{Z} \tag{1}$$

where $\mu \in \mathbb{R}$ and (ψ_j) satisfies $\sum_{j\in\mathbb{Z}} |\psi_j| < \infty$. X

Theorem

The series in Equation (1) *converges in* $\mathbb{L}^2$ *and the linear process* X *defined above is stationary. (see Proposition 3.1.2 in [BD13]).*

Exercise

Compute the mean and autocovariance functions of $(X_t)_{t\in\mathbb{Z}}$.

Examples of linear processes

Exercise

- Show that the following processes are particular linear processes
  - the white noise process
  - the MA(1) process.
- Consider a linear process as defined on slide 29, put $\mu = 0$,

$$\begin{cases} \psi_j = \phi^j & \text{if } j \geq 0 \\ \psi_j = 0 & \text{if } j < 0 \end{cases}$$

and suppose $|\phi| < 1$. Show that X is in fact an AR(1) process.

Estimation

Suppose that X is a stationary time series and recall that

$$\mu_X(t) = \mu, \quad \gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h}) \text{ and } \rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

for all $t, h \in Z$.

Estimation

From observations $X_1, \ldots, X_n$ (from the stationary time series X), we can compute

- the **sample mean** $\bar{X} = \frac{1}{n} \sum_{t=1}^{n} X_t$
- the **sample autocovariance function**

$$\hat{\gamma}_X(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X})(X_t - \bar{X}) \quad \forall -n < h < n$$

- the **sample autocorrelation function**

$$\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}.$$

Warning : $\gamma_X(h) = \mathrm{Cov}(X_t\,, X_{t+h})$ but the sample autocorrelation function is not the corresponding empirical covariance ! !

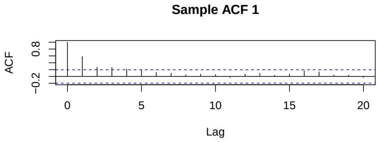
$$\frac{1}{n}\sum_{t=1}^{n-|h|}(X_{t+|h|} - \bar{X})(X_t - \bar{X}) \neq$$

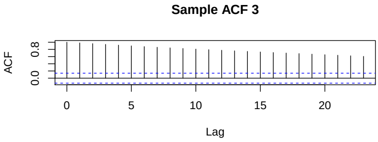
$$\frac{1}{n-|h|}\sum_{t=1}^{n-|h|}\Big(X_{t+|h|} - \frac{1}{n-|h|}\sum_{t=1}^{n-|h|}X_{t+h}\Big)\Big(X_t - \frac{1}{n-|h|}\sum_{t=1}^{n-|h|}X_t\Big)$$

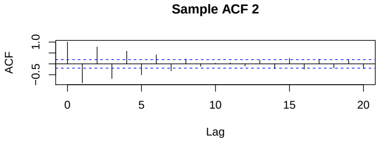
Examples of sample ACF

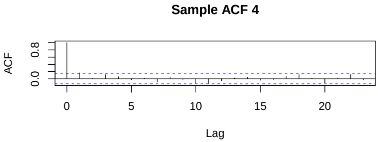
Exercise

Can you find the generating time series models (white noise, MA(1), AR(1), random noise with drift) associated with the sample ACF ?









Examples of sample ACF

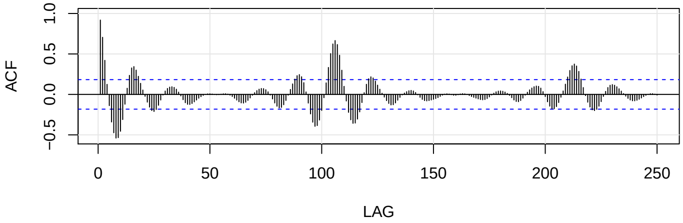


FIG.: The ACF of speech data example on slide 8

Notice :

- the regular repetition of short peaks with decreasing amplitude.

Properties of $\bar{X}_n$

Theorem

If X is a stationary time series, the sample mean verifies

$$\mathbb{E}(\bar{X}_n) = \mu$$
$$\mathbb{V}(\bar{X}_n) = \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma(h).$$

As a consequence, if

$$\sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty \quad \textit{then} \quad n\mathbb{V}(\bar{X}_n) \overset{n\to\infty}{\to} \sum_{h=-\infty}^{\infty}\gamma(h) = \sigma^2\sum_{h=-\infty}^{\infty}\rho(h)$$

and $\bar{X}_n$ converges in $\mathbb{L}^2$ to μ .

Notice that, in the independent case, $n\mathbb{V}(\bar{X}_n) \overset{n\to\infty}{\to} \sigma^2$. The correlation has hence the effect of reducing to sample size from n to $n/\sum_{h=-\infty}^{\infty}\rho(h)$.
See Appendix A [SS10]

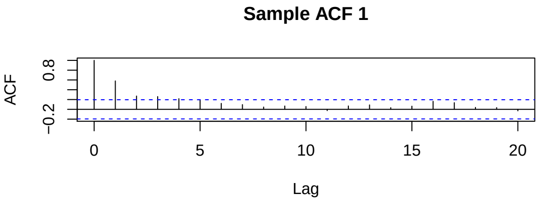
Large sample property

Theorem

Under general conditions, if X is a white noise, then for n large, the sample ACF, $\hat{\rho}_X(h)$, for $h = 1, 2, \ldots, H$, where H is fixed but arbitrary, is approximately normally distributed with zero mean and standard deviation given by

$$\sigma_{\hat{\rho}_X(h)} = \frac{1}{\sqrt{n}}$$

Consequence : only the peaks outside of $\pm 2/\sqrt{n}$ may be considered to be significant.



See Appendix A [SS10]

ACF and prediction

Linear predictor and ACF

Let X be a stationary time series with ACF ρ. The linear predictor $\hat{X}_{n+h}^{\{n\}}$ of X_{n+h} given X_n is defined as

$$\hat{X}_{n+h}^{\{n\}} = \underset{a,b}{\operatorname{argmin}}\, \mathbb{E}\Big(\big(X_{n+h} - (aX_n + b)\big)^2 \Big) = \rho(h)(X_n - \mu) + \mu$$

Exercise

Prove the result.

Notice that

- linear prediction needs only second order statistics, we'll see later that it is a crucial property for forecasting.
- the result extends to longer histories $(X_n, X_{n-1}, \dots)$.

Chapter 2 : Chasing stationarity, exploratory data analysis

- **Why do we need to chase stationarity ?**
  Because we want to do statistics : averaging lagged products over time, as in the previous section, has to be a sensible thing to do.
- **But....**
  Real time series are often non-stationary, so we need methods to "stationarize" the series.

An example I

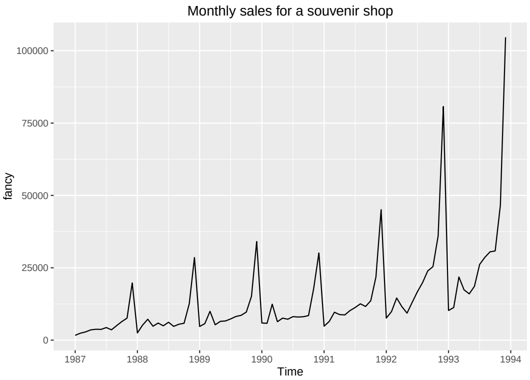


FIG. : Monthly sales for a souvenir shop on the wharf at a beach resort town in Queensland, Australia. [MWH08]

An example II

Notice that the variance grows with the mean, this usually calls for a log transformation ($X \to \log(X)$), which is part of the general family of Box-Cox transformation

$$\begin{cases} X \to X^{\lambda-1}/\lambda & \lambda \neq 0 \\ X \to \log(X) & \lambda = 0 \end{cases}$$

An example III

FIG.: Log of monthly sales.

The series is not yet stationary because there are a trend and a seasonal components.

An example IV

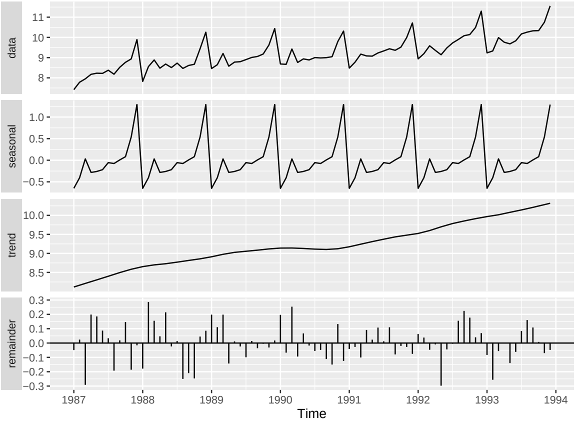


FIG.: Decomposition of monthly sales with `slt` function in R

Classical decomposition of a time series

$$Y_t = T_t + S_t + X_t$$

où

- $T = (T_t)_{t\in\mathbb{Z}}$ is the trend
- $S = (S_t)_{t\in\mathbb{Z}}$ is the seasonality
- $X = (X_t)_{t\in\mathbb{Z}}$ is a stationary centered time series.

Back to the global temperature I

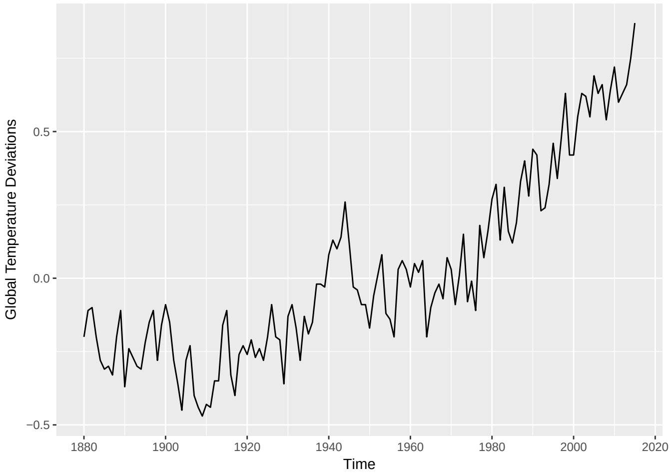


FIG. : Global temperature deviation (in $^{\circ}C$) from 1880 to 2015, with base period 1951-1980 - see slide 7

Back to the global temperature II

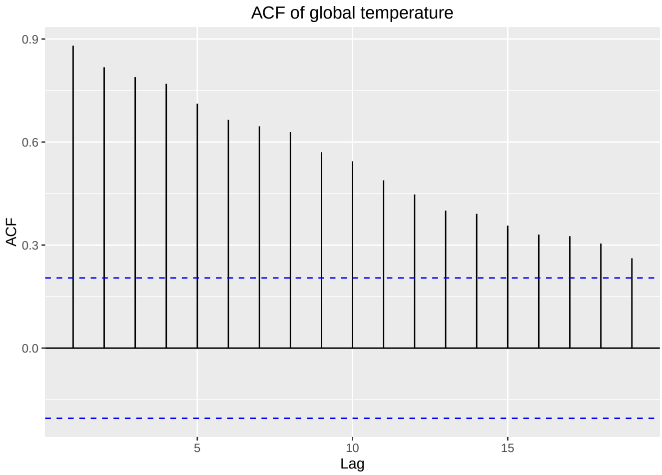


FIG.: ACF of global temperature deviation

Back to the global temperature III

We model this time series as

$$Y_t = T_t + X_t$$

and are now looking for a model for $T = (T_t)_{t\in\mathbb{Z}}$.
Looking at the series, two possible models for T are

- (model 1) a linear function of t $T_t = \beta_1 + \beta_2 t$
- (model 2) a random walk with drift $T_t = \delta + T_{t-1} + \eta_t$, where η is a white noise (see slide 19).

In both models, we notice that

$$Y_t - Y_{t-1} = T_t - T_{t-1} + \omega_t - \omega_{t-1} = \beta_2 + \omega_t - \omega_{t-1} \text{ (model 1)}$$
$$Y_t - Y_{t-1} = T_t - T_{t-1} + \omega_t - \omega_{t-1} = \delta + \eta_t + \omega_t - \omega_{t-1} \text{ (model 2)}$$

are stationary time series (check this fact as an exercise).

Back to the global temperature IV

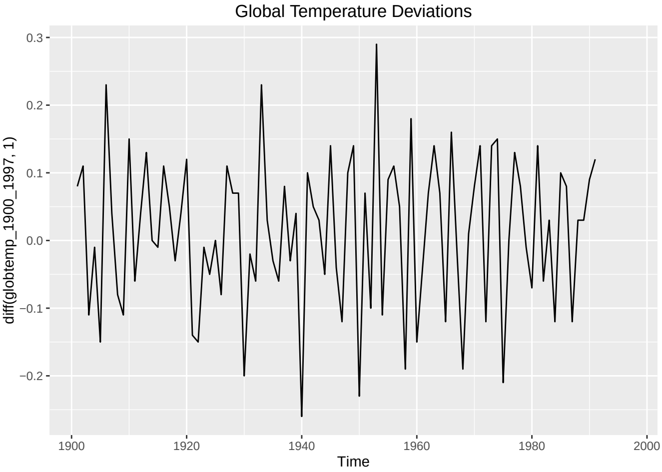


FIG.: Differenced global temperature deviation

Back to the global temperature V

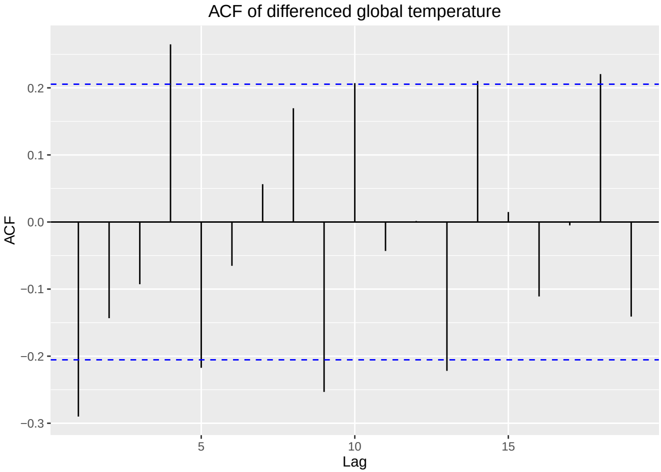


FIG.: Differenced global temperature deviation

Not far from a white noise!!

Backshift operator

Backshift operator

For a time series X, we define the **backshift operator** as

$$BX_t = X_{t-1},$$

similary

$$B^k X_t = X_{t-k}.$$

Difference of order d

Differences of order d are defined as

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

To stationarize the global temperature series, we applied the 1st order difference to it.
See http ://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html for an example of 2nd order integrated ts.

Moving average smoother

Moving average smoother

For a time series X,

$$M_t = \sum_{j=-k}^{k} a_j X_{t-j}$$

with $a_j = a_{-j} \geq 0$ and $\sum_{j=-k}^{k} a_j = 1$ is a symmetric moving average.

Note : `slt` function in R uses loess regression, the moving average smoother is just a loess regression with polynomials of order 1. More details on this on http ://www.wessa.net/download/stl.pdf, [CCT90].

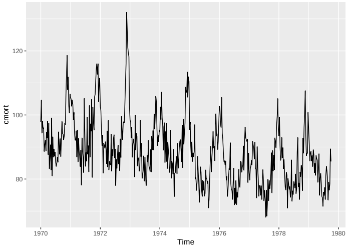


FIG. : Average daily cardiovascular mortality in Los Angeles county over the 10 year period 1970-1979.

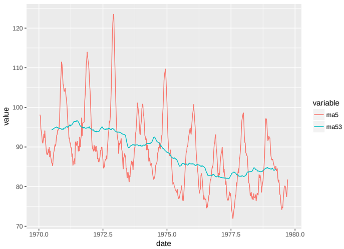


FIG.: Smoothed (ma 5 and 53) mortality

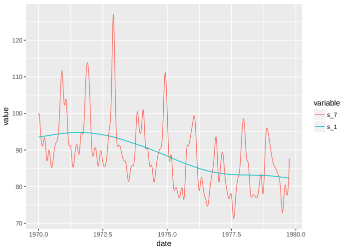


FIG.: Smoothed (splines) mortality

Chapter 3 : ARMA models

Introduction

We now consider that we have estimated the trend and seasonal components of

$$Y_t = T_t + S_t + X_t$$

Aim of the chapter : to propose to the time series X via ARMA models. They allow

- to describe this time series
- to forecast.

Key fact : we know that

- for every stationary process with autocovariance function γ verifying $\lim_{h \to \infty} \gamma(h) = 0$, it is possible to find an ARMA process with the same autocovariance function, see [BD13].
- The Wold decomposition (see [SS10] Appendix B) also plays an important role. Its says that every stationary process is the sum of a MA(∞) process and a deterministic process.

AR(1)

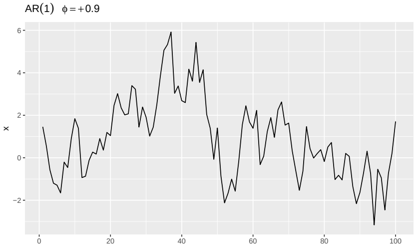
Exercice

Consider a time series X following the AR(1) model

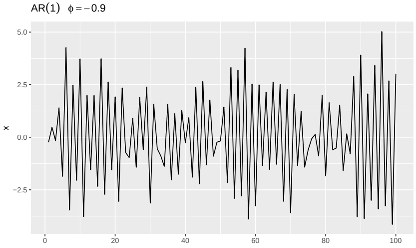
$$X_t = \phi X_{t-1} + \omega_t \; \forall t \in \mathbb{Z}.$$

1. Show that for all $k > 0$ $X_t = \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j \omega_{t-j}$.
2. Assume that $|\phi| < 1$ and prove $X_t \overset{\mathbb{L}^2}{=} \sum_{j=0}^{\infty} \phi^j \omega_{t-j}$.
3. Assume now that $|\phi| > 1$ and prove that
   3.1 $\sum_{j=0}^{k-1} \phi^j \omega_{t-j}$ does not converge in $\mathbb{L}^2$
   3.2 one can write $X_t = -\sum_{j=1}^{\infty} \phi^{-j} w_{t+j}$
   3.3 Discuss why the case $|\phi| > 1$ is useless.

The case where $|\phi| = 1$ is a random walk (slide 19) and we already proved that this is not a stationary time series.



AR(1) ϕ = +0.9
6
4
2
0
−2
x
0
20
40
60
80
100



AR(1) ϕ = −0.9
5.0
2.5
0.0
−2.5
x
0
20
40
60
80
100

Note on polynomials

Notice that manipulating operators like $\phi(\mathrm{B})$ is like manipulating polynomials with complex variables.
In particular :

$$\frac{1}{1-\phi z} = 1 + \phi z + \phi^2 z + \dots$$

provided that $|\phi| < 1$ and $|z| \leq 1$.

Causality

Causal linear process

A linear process X is said to be **causal** when there is

- a power series $\pi : \pi(x) = \pi_0 + \pi_1 x + \pi_2 x^2 , \ldots,$
- with $\sum_{j=0}^{\infty} |\pi_j| < \infty$
- and $X_t = \pi(B)\,\omega_t$

ω is a white noise $WN(0\,,\sigma^2)$.
In this case X_t is $\sigma\{\,\omega_t\,,\omega_{t-1}\,,\ldots\}$-measurable.

We will exclude non-causal AR models from consideration. In fact this is not a restriction because we can find causal counterpart to such process.

Exercise

Consider the non-causal AR(1) model $X_t = \phi X_{t-1} + \omega_t$ with $|\phi| > 1$ and suppose that $\omega \sim i.i.d\, \mathcal{N}(0\,,\sigma^2)$

1. Which distribution has X_t ?
2. Define the time series $Y_t = \phi^{-1} Y_{t-1} + \eta_t$ with $\eta \sim i.i.d\, \mathcal{N}(0\,,\sigma^2/\phi^2)$. Prove that X_t and Y_t have the same distribution.

Autoregressive model

AR(p)

An autoregressive model of order p is of the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \omega_t \ \forall t \in \mathbb{Z}$$

where X is assumed to be stationary and ω is a white noise $WN(0, \sigma^2)$. We will write more concisely

$$\Phi(B) X_t = \omega_t \ \forall t \in \mathbb{Z}$$

where ϕ is the polynomial of degree p $\phi(x) = (1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p)$.

Without loss of generality, we assume that each X_t is centered.

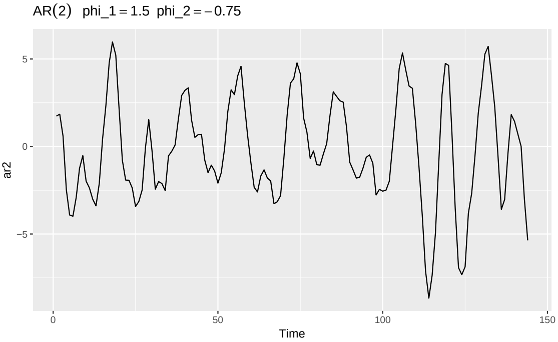
Condition of existence and causality of AR(p)

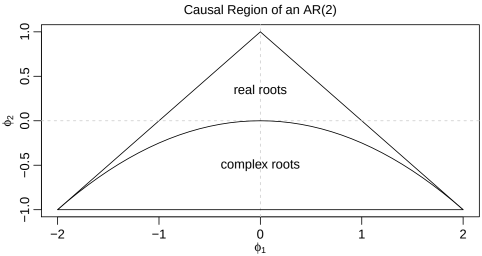
A stationary solution to $\Phi(B)X_t = \omega_t \; \forall t \in \mathbb{Z}$ exists if and only if

$$\phi(z) = 0 \Longrightarrow |z| \neq 1.$$

In this case, this defines an AR(p) process, which is causal iff in addition

$$\phi(z) = 0 \Longrightarrow |z| > 1.$$


AR(2) phi_1 = 1.5 phi_2 = −0.75
ar2
5
0
−5
0
50
100
150
Time


Causal Region of an AR(2)
real roots
complex roots
φ2
1.0
0.5
0.0
−0.5
−1.0
−2
−1
0
1
2
φ1

Moving average model

MA(q)

An moving average model of order q is of the form

$$X_t = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \ldots + \theta_q \omega_{t-q} \; \forall t \in \mathbb{Z}$$

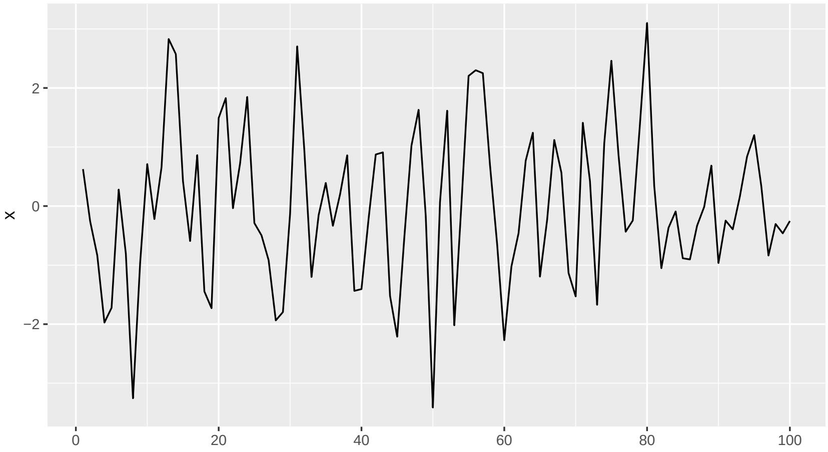
where ω is a white noise $WN(0\,,\sigma^2)$. We will write more concisely

$$X_t = \Theta(B)\omega_t \; \forall t \in \mathbb{Z}$$

where θ is the polynomial of degree q $\theta(x) = (1 - \theta_1 x - \theta_2 x^2 - \ldots - \theta_q x^q)$.

Unlike the AR model, the **MA model is stationary for any values of the thetas.**

MA(1) $\theta = +0.9$



Invertibility I

Invertibility of a MA(1) process

Consider the MA(1) process

$$X_t = \omega_t + \theta\omega_{t-1} = (1+\theta B)\omega_t \ \forall t \in \mathbb{Z}$$

where ω is a white noise $WN(0\,,\sigma^2)$.
Show that

- If $|\theta| < 1,\ \omega_t = \sum_{j=0}^{\infty}(-\theta)^j X_{t-j}$
- If $|\theta| > 1,\ \omega_t = -\sum_{j=1}^{\infty}(-\theta)^{-j} X_{t+j}$

In the first case, X is invertible.

Invertibility

A linear process X is **invertible** when there is

- a power series $\pi : \pi(x) = \pi_0 + \pi_1 x + \pi_2 x^2\,, \ldots,$
- with $\sum_{j=0}^{\infty} |\pi_j| < \infty$
- and $\omega_t = \pi(B)X_t$

ω is a white noise $WN(0\,,\sigma^2)$.

Invertibility II

Exercise

Consider the non-invertible MA(1) model $X_t = \omega_t + \theta\,\omega_{t-1}$ with $|\theta| > 1$ and suppose that $\omega \sim i.i.d.\mathcal{N}(0, \sigma^2)$

1. Which distribution has X_t?
2. Can we define an invertible time series Y defined through a new Gaussian white noise η such that X_t and Y_t have the same distribution $(\forall t)$?

Autoregressive moving average model

Autoregressive moving average model ARMA(p, q)

An ARMA(p, q) process X is a stationary process that is defined through

$$\Phi(B)X_t = \Theta(B)\omega_t$$

where $\omega \sim WN(0, \sigma^2)$, Φ is a polynomial of order p, Θ is a polynomial of order q and Φ and Θ have no common factors.

Exercise

Consider the process X defined by $X_t - 0.5X_{t-1} = \omega_t - 0.5\,\omega_{t-1}$. Is it trully an ARMA(1,1) process ?

Stationarity, causality and invertibility

Theorem

Consider the equation $\Phi(B)X_t = \Theta(B)\,\omega_t$, where Φ and Θ have no common factors.

- There exists a stationary solution iff

$$\phi(z) = 0 \Leftrightarrow |z| \neq 1.$$

- This process ARMA(p, q) is causal iff

$$\phi(z) = 0 \Leftrightarrow |z| > 1.$$

- It is invertible iff the roots of $\theta(z)$ are outside the unit circle.

Exercise

Discuss the stationarity, causality and invertibility of $(1 - 1.5B)X_t = (1 + 0.2B)\,\omega_t$.

Theorem

Let X be an ARMA process defined by $\Phi(B)X_t = \Theta(B)\,\omega_t$.
If

$$\forall |z| = 1 \ \ \theta(z) \neq 0,$$

then there are polynomials $\tilde{\phi}$ and $\tilde{\theta}$ and a white noise sequence $\tilde{\omega}$ such that X satisfies

- $\tilde{\Phi}(B)X_t = \tilde{\Theta}(B)\tilde{\omega}_t$,
- and is a causal,
- invertible ARMA process.

We can now **consider only causal and invertible ARMA processes.**

The linear process representation of an ARMA

Causal and invertible representations

Consider a causal, invertible ARMA process defined by $\Phi(B)X_t = \Theta(B)\,\omega_t$. It can be rewritten

- as a MA(∞) :

$$X_t = \frac{\Theta(B)}{\Phi(B)}\omega_t = \psi(B)\omega_t = \sum_{k\geq 0}\psi_k\omega_{t-k}$$

- or as an AR((∞))

$$\omega_t = \frac{\Phi(B)}{\Theta(B)}X_t = \pi(B)X_t = \sum_{k\geq 0}\pi_k X_{t-k}$$

Notice that both π_0 and ψ_0 equal 1 and (ψ_k) and (π_k) are entirely determined by (ϕ_k) and (θ_k).

Autocovariance function of an ARMA

Autocovariance of an ARMA

The autocovariance function of an ARMA(p, q) follows from its MA(∞) representation and equals

$$\gamma(h) = \sigma^2 \sum_{k \geq 0} \psi_k \psi_{k+h} \; \forall h \geq 0.$$

Exercise

- Compute the ACF of a causal ARMA$(1, 1)$.
- Show that the ACF of this ARMA verifies a linear difference equation of order 1. Solve this equation.
- Compute ϕ and θ from the ACF.

Chapter 4 : Linear prediction and partial autocorrelation function

Introduction

We'll see that if we know

- the **orders (*p* and *q*)** and
- the **coefficients**

of the ARMA model under consideration, we can build predictions and prediction intervals.

Just to be sure....

- The linear space $\mathbb{L}^2$ of r.v. with finite variance with the inner-product $\langle X, Y\rangle = \mathbb{E}(XY)$ is an Hilbert space.
- Now considering a time series X with $X_t \in \mathbb{L}^2$ for all t
  - the subspace $\mathcal{H}_n = \text{span}(X_1, \dots, X_n)$ is a closed subspace of $\mathbb{L}^2$ hence
  - for all $Y \in \mathbb{L}^2$ there exists an unique projection $P(Y)$ in $\mathcal{H}_n$ such that, for all $\forall w \in \mathcal{H}_n$

$$\|P(Y) - Y\| \leq \|w - Y\|$$
$$\langle P(Y) - Y, w\rangle = 0.$$

Best linear predictor

Given $X_1, X_2, \ldots, X_n$, the **best linear m-step-ahead predictor** of X_{n+m} defined as

$$X_{n+m}^{(n)} = \alpha_0 + \phi_{n1}^{(m)} X_n + \phi_{n2}^{(m)} X_{n-1} + \phi_{nn}^{(m)} X_1 = \alpha_0 + \sum_{j=1}^{n} \phi_{nj}^{(m)} X_{n+1-j}$$

is the orthogonal projection of X_{n+m} onto span$\{1, X_1, \ldots, X_n\}$. In particular, it satisfies the **prediction equations**

$$\mathbb{E}(X_{n+m}^{(n)} - X_{n+m}) = 0$$
$$\mathbb{E}((X_{n+m}^{(n)} - X_{n+m})X_k) = 0 \;\; \forall k = 1, \ldots, n$$

We'll now compute α_0 and the $\phi_{nj}^{(m)}$'s.

Derivation of α_0

We get

$$X_{n+m}^{(n)} - \mu = \alpha_0 + \sum_{j=1}^{n} \phi_{nj}^{(m)} X_{n+1-j} - \mu = \sum_{j=1}^{n} \phi_{nj}^{(m)} (X_{n+1-j} - \mu)$$

- Thus, we'll ignore α_0 and put $\mu = 0$ until we discuss estimation.
- There are two consequences
  1. the projection of X_{n+m} on onto span$\{1, X_1, \dots, X_n\}$ is in fact the projection onto span$\{X_1, \dots, X_n\}$
  2. $\mathbb{E}(X_k X_l) = \text{Cov}(X_k, X_l)$

Derivation of the $\phi_{nj}^{(m)}$'s

As $X_{n+m}^{(n)}$ satisfies the prediction equations of slide 74, we can write for all $k = 1, \ldots, n$

$$
\begin{aligned}
& \mathbb{E}((X_{n+m}^{(n)} - X_{n+m})X_k) = 0 \\
\iff & \sum_{j=1}^{n} \phi_{nj}^{(m)} \mathbb{E}\big(X_{n+1-j}X_{n+1-k}\big) = \mathbb{E}\big(X_{n+m}X_{n+1-k})\big) \\
\iff & \sum_{j=1}^{n} \alpha_j \gamma(k-j) = \gamma(m+k-1)
\end{aligned}
$$

This can rewritten in matrix notation.

Prediction

Prediction equations

The $\phi_{nj}^{(m)}$'s verify

$$\Gamma_n \phi_n^{(m)} = \gamma_n^{(m)}$$

where

$$\Gamma_n = \Big(\gamma(k-j)\Big)_{1 \le j,k \le n}$$

$$\phi_n^{(m)} = \left(\phi_{n1}^{(m)}, \dots, \phi_{nn}^{(m)}\right)^\top,$$

$$\gamma_n^{(m)} = \left(\gamma(m), \dots, \gamma(m+n-1)\right)^\top.$$

Prediction error

The mean square prediction error is given by

$$P_{n+m}^n = \mathbb{E}\Big(\big(X_{n+m} - X_{n+m}^{(n)}\big)^2\Big) = \gamma(0) - (\gamma_n^{(m)})^\top \Gamma_n^{-1} \gamma_n^{(m)}.$$

Forecasting an AR(2)

Exercise

Consider the causal AR(2) model $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \omega_t$.

1. Determine the one-step-ahead $X_3^{(2)}$ prediction of X_3 based on X_1 , X_2 from the prediction equations.
2. From causality, determine $X_3^{(2)}$.
3. How $\phi_{21}^{(1)}$, $\phi_{22}^{(1)}$ and ϕ_1 , ϕ_2 are related ?

PACF

Partial autocorrelation function

The **partial autocorrelation function** (PACF) of a stationary time series X is defined as

$$\phi_{11} = \text{cor}\big(X_1, X_0\big) = \rho(1)$$
$$\phi_{hh} = \text{cor}\big(X_h - X_h^{(h-1)}, X_0 - \hat{X}_0^{(h-1)}\big) \text{ for } h \geq 2,$$

where $\hat{X}_0^{(h-1)}$ is the orthogonal projection of X_0 onto span$\{X_1, \ldots, X_{h-1}\}$.

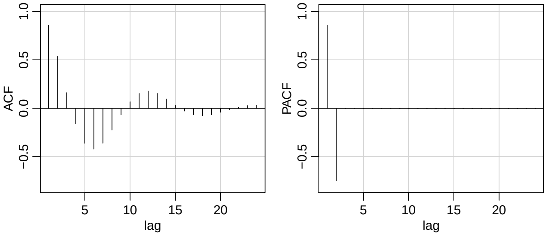
Notice that

- $X_h - X_h^{(h-1)}$ and $X_0 - \hat{X}_0^{(h-1)}$ are, by construction, uncorrelated with $\{X_1, \ldots, X_{h-1}\}$, so ϕ_{hh} is the correlation between X_h and X_0 with the linear dependence of $X_1, \ldots, X_{h-1}$ on each removed.
- The coefficient ϕ_{hh} is also the last coefficient (i.e. $\phi_{hh}^{(1)}$) in the best linear one-step-ahead prediction of X_{h+1} given $X_1, \ldots, X_h$.

PACF of an AR(p) model

Consider the causal AR(p) model $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \omega_t$

1. Consider $p = 2$ and verify that $X_{n+1}^{(n)} = \phi_1 X_n + \phi_2 X_{n-1}$. Deduce the value of the PACF for $h > 2$
2. In the general case, deduce the value of the PACF for $h > p$.



PACF of invertible MA models

Exercise : PACF of a MA(1) model

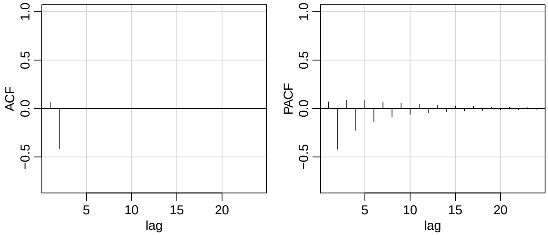
Consider the invertible MA(1) model $X_t = \omega_t + \theta \omega_{t-1}$

1. Compute $\hat{X}_3^{(2)}$ and $\hat{X}_1^{(2)}$, the orthogonal projections of X_3 and X_1 onto span$\{X_2\}$.
2. Deduce the first two values of the PACF.

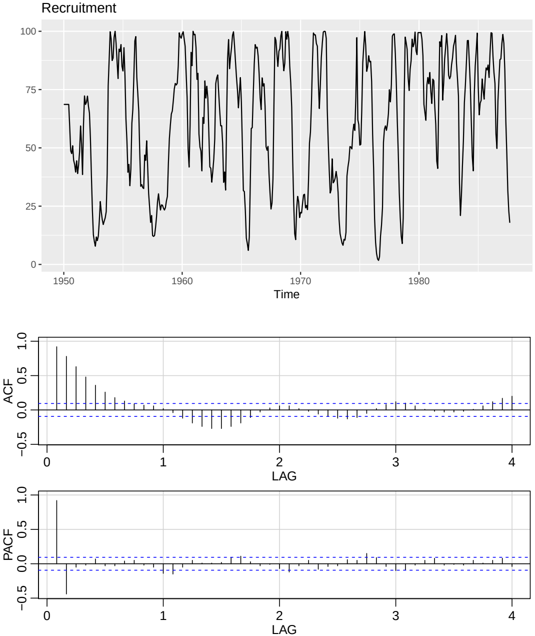
More calculations (see Problem 3.23 in [BD13]) give

$$\phi_{hh} = -\frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2(h+1)}}.$$

In general, the PACF of a MA(q) model does not vanish for larger lag, it is however bounded by a geometrically decreasing function.


ACF
1.0
0.5
0.0
−0.5
5
10
15
20
lag
PACF
1.0
0.5
0.0
−0.5
5
10
15
20
lag

An AR(2) model for the recruitement series



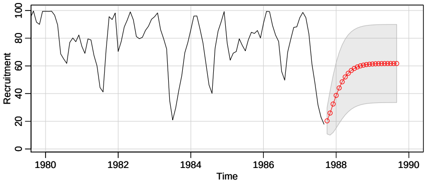


FIG.: Twenty-four month forecats for the Recruitment series shown on slide 11

ACF and PACF

So far, we show that

Model	**ACF**	**PACF**
AR(p)	decays	zero for $h > p$
MA(q)	zero for $h > q$	decays
ARMA	decays	decays

- We can use these results to **build a model**.
- And we know how to **forecast in an AR(p) model**.
- It remains to give algorithms that will allow to forecast in MA and ARMA models.

Innovations

So far, we have written X_{n+1}^n as $\sum_{j=1}^n \phi_{nj}^{(m)} X_{n+1-j}$ i.e. as the projection of X_{n+1} onto span$\{X_1, \ldots, X_n\}$ but we clearly have

$$\text{span}\{X_1, X_2 - X_2^1, X_3 - X_3^2, \ldots, X_n - X_n^{n-1}\}.$$

Innovations

The values $X_t - X_t^{t-1}$ are called the **innovations**. They verify $X_t - X_t^{t-1}$ is orthogonal to $\mathtt{span}\{X_1, \ldots, X_{t-1}\}$.

As a consequence, we can rewrite

$$X_{n+1}^n = \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - X_{n+1-j}^{n-j})$$

The one-step-ahead predictors X_{t+1}^t and their mean-squared errors P_{t+1}^t can be calculated iteratively via the innovations algorithm.

The innovations algorithm

The one-step-ahead predictors can be iteratively be computed via

$$X_1^0 = 0, P_1^0 = \gamma(0) \text{ and } t = 1, 2, \dots$$

$$X_{t+1}^t = \sum_{j=1}^{t} \theta_{tj}(X_{t+1-j} - X_{t+1-j}^{t-j})$$

$$P_{t+1}^t = \gamma(0) - \sum_{j=0}^{t-1} \theta_{t,t-j}^2 P_{j+1}^j \text{ where}$$

$$\theta_{t,t-h} = \left(\gamma(t-h) - \sum_{k=0}^{h-1} \theta_{h,h-k}\theta_{t,t-k}P_{k+1}^k\right)(P_{h+1}^h)^{-1} \quad h = 0, 1, \dots, t-1$$

This can be solve by calculting P_1^0, θ_{11}, P_2^1, θ_{22}, θ_{21}, etc.

Prediction for an MA(1)

Exercise

Consider the MA(1) model $X_t = \omega_t + \theta\,\omega_{t-1}$ with $\omega \sim WN(0\,,\sigma^2)$. We know that $\gamma(0) = \sigma^2(1+\theta^2)$, $\gamma(1) = \theta\sigma^2$ and $\gamma(h) = 0$ for $h \geq 2$.

Show that

$$X_{n+1}^n = \theta\frac{X_n - X_n^{n-1}}{r_n}$$

with

$$r_n = P_n^{n-1}/\sigma^2.$$

The innovations algorithm for the ARMA(p, q) model

Consider an ARMA(p, q) model

$$\Phi(B)X_t = \Theta(B)\omega_t \text{ with } \omega \sim WN(0, \sigma^2).$$

Let $m = \max(p, q)$, to simplify calculations, the innovation algorithm is not applied directly to X but to

$$\begin{cases} W_t = \sigma^{-1}X_t & t = 1, \ldots, m \\ W_t = \sigma^{-1}\Phi(B)X_t & t > m. \end{cases}$$

see page 175 of [BD13]

Infinite past

We will now show that it is easier for a **causal, invertible** ARMA process

$$\Phi(B)X_t = \Theta(B)\omega_t$$

to approximate $X_{n+h}^{(n)}$ by a **truncation** of the projection of X_{n+h} onto the infinite past

$$\bar{\mathcal{H}}_n = \bar{\text{span}}\{X_n, X_{n-1}, \ldots\} = \bar{\text{span}}\{X_k, k \leq n\}.$$

The projection onto $\bar{\mathcal{H}}_n = \bar{\text{span}}(X_k\,, k \leq n)$ can be defined as

$$\lim_{k\to\infty} P_{\text{span}(X_{n-k},\ldots,X_n)}$$

We will define

$$\tilde{X}_{n+h} \text{ and } \tilde{\omega}_{n+h}$$

as the projections of X_{n+h} and ω_{n+h} onto $\bar{\mathcal{H}}_n$.

Causal and invertible

Recall (see slide 69) that since X is causal and invertible, we may write

- $X_{n+h} = \sum_{k \geq 0} \psi_k \, \omega_{n+h-k}$ (MA(∞) representation)
- $\omega_{n+h} = \sum_{k \geq 0} \pi_k X_{n+h-k}$ (AR((∞)) representation).

Now, applying the projection operator onto $\mathcal{M}_n$ on both sides of both equations, we get

$$\tilde{X}_{n+h} = \sum_{k \geq 0} \psi_k \tilde{\omega}_{n+h-k} \tag{2}$$

$$\tilde{\omega}_{n+h} = \sum_{k \geq 0} \pi_k \tilde{X}_{n+h-k}. \tag{3}$$

Iteration

We get

$$\tilde{X}_{n+h} = -\sum_{k\geq 1} \pi_k \tilde{X}_{n+h-k} \text{ and}$$

$$\mathbb{E}\big((X_{n+h} - \tilde{X}_{n+h})^2\big) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 .$$

As $\tilde{X}_t = X_t$ for all $t \leq n$, we can define recursively

$$\tilde{X}_{n+1} = -\sum_{k\geq 1} \pi_k X_{n+h-k}$$

$$\tilde{X}_{n+2} = -\pi_1 \tilde{X}_{n+1} - \sum_{k\geq 2} \pi_k X_{n+h-k}$$

$$\dots .$$

Truncation

In practice, we do not observe the past from $-\infty$ but only $X_1, \ldots, X_n$, but we can use a truncated version

$$\tilde{X}_{n+1}^{T} = -\sum_{k=1}^{n} \pi_k X_{n+h-k}$$

$$\tilde{X}_{n+2}^{T} = -\pi_1 \tilde{X}_{n+1}^{T} - \sum_{k=2}^{n+1} \pi_k X_{n+h-k}$$

$$\ldots$$

and $\mathbb{E}\big((X_{n+h} - \tilde{X}_{n+h})^2\big) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$ is used an approximation of the predictor error.

Chapter 5 : Estimation and model selection

Introduction

We saw in the last chapter, that if we know

- the **orders (p and q)** and
- the **coefficients**

of the ARMA model under consideration, we can build predictions and prediction intervals.

The aim of this chapter is to present

- methods for **estimating the coefficients** when the orders (p and q) are known
- **model selection methods**, i.e. methods for selecting p and q

Caution :

- To avoid confusion, **true parameters now wear a star** :

$$\sigma^{2,\star}, \phi_1^\star, \ldots, \phi_p^\star, \theta_1^\star, \ldots, \theta_q^\star$$

- we have a sample $(X_1, \ldots, X_n)$ to build estimators.

Moment estimations

We assume that $\mu^\star = 0$ (without loss of generality) in Chapter 4. We now consider causal and invertible ARMA processes of the form

$$\Phi(B)(X_t - \mu^\star) = \Theta(B)\omega_t$$

where $\mathbb{E}(X_t) = \mu^\star$

Estimation of the mean

For a stationary time series, the moment estimator of $\mu^\star$ is the sample mean $\bar{X}_n$.

AR(1) model

Give the moment estimators in a stationary AR(1) model.

Moment estimators for AR(p) models

Yule-Walker equations for an AR(p)

The autocovariance function and parameters of the AR(p) model verify

$$\Gamma_p \phi^\star = \gamma_p \quad \text{and} \quad \sigma^{2,\star} = \gamma(0) - (\phi^\star)^\top \gamma_p$$

where

$$\Gamma_p = \Big(\gamma(k-j)\Big)_{1 \le j,k \le p} \quad \phi^\star \quad = (\phi_1^\star, \dots, \phi_p^\star)^\top, \text{ and } \gamma_p = \big(\gamma(1), \dots, \gamma(p)\big)^\top.$$

This leads to

$$\widehat{\phi} = \widehat{\Gamma}_p^{-1} \widehat{\gamma}_p \quad \text{and} \quad \hat{\sigma}^2 = \hat{\gamma}(0) - \widehat{\phi}^\top \widehat{\gamma}_p.$$

AR(2)

Verify the Yule Walker equation for a causal AR(2) model.

Asymptotics

The only case in which the moment method is (asymptotically) efficient is the AR(p) model.

Asymptotic distribution of moment estimators

Under mild conditions on ω, and if the AR(p) is causal, the Yule-Walker estimators verify

$$\sqrt{n}(\widehat{\phi} - \phi^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^{2,\star}\Gamma_p^{-1})$$
$$\hat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^{2,\star}.$$

Likelihood of an causal AR(1) model I

We now deal with maximum likelihood estimation, we assume that

$$\omega \sim i.i.d. \mathcal{N}(0, \sigma^{2,\star}).$$

The likelihood the causal AR(1) model

$$X_t = \mu^\star + \phi^\star (X_{t-1} - \mu^\star) + \omega_t$$

is given by

$$\mathcal{L}_n(\mu, \phi, \sigma^2) = f_{\mu,\phi,\sigma^2}(X_1, \ldots, X_n)$$
$$= f_{\mu,\phi,\sigma^2}(X_1) f_{\mu,\phi,\sigma^2}(X_2|X_1) f_{\mu,\phi,\sigma^2}(X_3|X_1, X_2) \ldots f_{\mu,\phi,\sigma^2}(X_n|X_1, X_2, \ldots, X_{n-1})$$

Likelihood of an causal AR(1) model II

We can now write the log-likelihood

$$\begin{aligned}\ell_n(\mu, \phi, \sigma^2) &= \log \mathcal{L}_n(\mu, \phi, \sigma^2) \\ &= -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}S(\mu,\ \phi)\end{aligned}$$

with

$$S(\mu,\ \phi) = (1-\phi^2)(X_1-\mu) + \sum_{k=2}^{n}\big(X_k - \mu + \phi(X_{k-1}-\mu)\big)^2.$$

It is straightforward to see that

$$\hat{\sigma}^2 = \frac{1}{n}S(\hat{\mu},\ \hat{\phi})$$

where

$$\hat{\mu}, \hat{\phi} = \text{argmin}_{\mu,\phi} \log(S(\mu,\ \phi)/n) - \frac{1}{n}\log(1-\phi^2).$$

Likelihood for causal, invertible ARMA model I

Consider the causal and invertible ARMA(p, q)

$$\Phi(B)X_t = \Theta(B)\omega_t,$$

when $\omega \sim i.i.d.\mathcal{N}(0, \sigma^{2,\star})$, one can show that

$$X_t|X_1, \ldots, X_{t-1} \sim \mathcal{N}(X_t^{(t-1)}, P_t^{t-1}) \quad \text{with}$$

$$P_t^{t-1} = \sigma^{2,\star} \sum_{j\geq 0} \psi_j^{2,\star} \prod_{k=1}^{t-1} (1 - \phi_{kk}^{2,\star}) := \sigma^{2,\star} r_t$$

see the details on pages 126 and following of [SS10] and the Durbin-Levinson algorithm (see page 112).

Likelihood for causal, invertible ARMA model II

Log-likelihood of an Gaussian ARMA(p, q) process

Denoting by β the vector $(\mu, \phi_1, \phi_p, \theta_1, \ldots, \theta_q)$, we have

$$\begin{aligned}\ell_n(\beta, \sigma^2) &= \log \mathcal{L}_n(\beta, \sigma^2) \\ &= -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\sum_{k=1}^{n}\log(r_k(\beta)) - \frac{1}{2\sigma^2}S(\beta)\end{aligned}$$

with

$$S(\beta) = \sum_{k=1}^{n}\left(\frac{X_k - X_k^{k-1}}{r_k(\beta)}\right)^2 \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}S(\widehat{\beta})$$

where

$$\widehat{\beta} = \text{argmin}_\beta \log(S(\beta)/n) - \frac{1}{n}\sum_{k=1}^{n}\log(r_k(\beta)).$$

The minimization problem is usually solve via Newton-Raphson algorithm.

Asymptotics

Asymptotic distribution of maximum likelihood estimators

Under appropriate conditions, and if the ARMA(p,q) is causal and invertible, the maximum likelihood estimators verify

$$\sqrt{n}(\widehat{\beta} - \beta^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^{2,\star}\Gamma_{p,q}^{-1,\star})$$
$$\hat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^{2,\star}$$

where the matrix $\Gamma_{p,q}^\star$ depends on $(\phi_1^\star, \dots, \phi_p^\star, \theta_1^\star, \dots, \theta_q^\star)$.

Other options involve in particular conditional sum of squares and the Gauss-Newton algorithm (details may be found on page 129 and following in [SS10]).

Model selection

Once the likelihood is given, model selection for the choice of parameters p and q can be performed via usual criteria.

To avoid confusion, we now denote by

$$\widehat{\beta}_{p,q} = (\hat{\mu}, \hat{\phi}_1, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q) \quad \text{and} \quad \hat{\sigma}^2_{p,q}$$

the maximum likelihood estimators in the ARMA(p, q) model, that is

$$\widehat{\beta}_{p,q}, \hat{\sigma}^2_{p,q} = \text{argmin}_{\beta_{p,q},\sigma^2} - 2\ell_n(\beta_{p,q}, \sigma^2)$$

AICc and BIC

AICc

The corrected AIC (Akaike Information Criterion) choose p and q that minimize

$$-2\ell_n(\widehat{\beta}_{p,q}, \hat{\sigma}^2) + 2\frac{(p+q+1)n}{n-p-q-2}.$$

BIC

The BIC (Bayesian Information Critetion) choose p and q that minimize

$$-2\ell_n(\widehat{\beta}_{p,q}, \hat{\sigma}^2) + \log(n)(p+q+1).$$

Residuals

The final step of model building is diagnostics on residuals.

Standardized innovations (residuals)

In a given model, the standardized innovations are given by

$$\frac{X_i - X_i^{(i-1)}}{\sqrt{P_i^{(i-1)}}}$$

for $i = 1, \ldots, n$.

Diagnostics on residuals

In the model is correct standardized innovations should behave like a white noise (even a Gaussian white noise if Gaussian maximum likelihood has been used.)

1. Plot the standardized innovations and their ACFs.
2. To check for normality, plot a histogram or a QQ-plot.
3. Verify that the ACF coefficients stay in the confidence interval for $h \geq 1$
4. Use a Ljung-Box test

Ljung-Box test

To test $\mathcal{H}_0 : \omega$ is a white noise, use the test statistic

$$Q = n(n+2) \sum_{h=1}^{H} \frac{\hat{\rho}^2_{p,q}(h)}{n-h}$$

where $\hat{\rho}_{p,q}$ is the sample ACF of the residuals in a given ARMA(p, q) model.

Q is asymptotically (under mild conditions) of χ^2_{H-p-q} distribution.

This is my link

Chapter 6 : Non-stationarity and seasonality

Integrated models

We now introduce a new class of models, which, based on ARMA models, incorporates a wide range of **non-stationary series**.

ARIMA(p,d,q) model

A process X is said to be ARIMA(p, d, q) if

$$(1-B)^d X_t$$

is an ARMA(p,q). We can rewrite

$$\Phi(B)(1-B)^d X_t = \Theta(B)\omega_t.$$

If $\mathbb{E}\big((1-B)^d X_t\big) = \mu$, we write

$$\Phi(B)(1-B)^d X_t = \delta + \Theta(B)\omega_t,$$

where $\delta = \mu(1-\phi_1-\phi_2-\ldots-\phi_p)$.

Caution : X is not stationary but $(1-B)^d X$ is, provided that the roots of Φ are outside the unit circle.

A special example

Random walk with drift

Consider the model of slide 19

$$X_t = \delta + X_{t-1} + \omega_t$$

with $X_0 = 0$

1. Check that X is an ARIMA(0,1,0).
2. Given data $X_1, \dots, X_n$, give the one-step-ahead prediction of X_{n+1}
3. Deduce the m-step-ahead prediction $X_{n+m}^{(n)}$
4. Compute the prediction error $P_{n+m}^{(n)}$.

Forecasting in ARIMA models

Exponentially weighted moving averages

Consider the process :

$$X_t = X_{t-1} + \omega_t - \theta\omega_{t-1},$$

with $|\theta| < 1$

1. Write it as an ARIMA(0,1,1).
2. Define $Y_t = \omega_t - \theta\,\omega_{t-1}$ and verify that Y is invertible.
3. Deduce that

$$X_t = \sum_{j=1}^{\infty}(1-\theta)\theta^{j-1}X_{t-j} + \omega_t$$

4. and finally that, based on the data $X_1, \dots, X_n$

$$X_{n+1}^{(n)} = (1-\theta)X_n + \theta X_n^{(n-1)}.$$

Building ARIMA model I

Here a the steps you should follow to build an ARIMA model

1. Construct a time plot of the data and inspect the graph for anomalies. For example, if the variability in the data depends upon time, you need to stabilize the variance via a Box-Cox transformation
2. Transform the data and construct a new time plot.
3. Choice of d : a look at the new time plot will help you determine if a differentiation is needed. If it is the case
   - Differentiate the series and inspect the time plot
   - If additional differentiating is required, apply the operator $(1 - B)^2$ and so on
   - Do not forget that a ACF decreasing too slowly is also a sign of non-stationarity
   - Caution : do not differentiate too many times

Counter example

Show that if ω_t is a white noise, $\omega_t - \omega_{t-1}$ is a MA(1) ! !

Building ARIMA model II

4. The next step is to identify reasonable values (or a set of reasonable values) for q, p
   4.1 Represent the ACFs and PACFs of the differentiated series (they can be more than one if you hesitate between two values for d) and
   4.2 choose few reasonable values for q and p
5. At this stage, you should have few preliminary reasonable values for d, q and p. Estimate the parameters in the different models and compute their AICc and BIC.
6. Choose one model and conduct diagnostic tests on its residuals

Back to seasonality

Can you you think of a model for data with these sample ACF and PACF ?

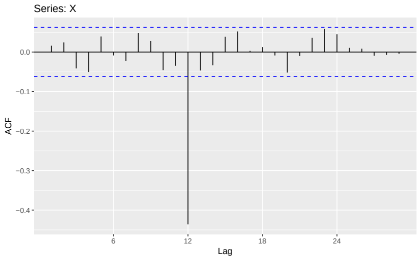


FIG.: ACF

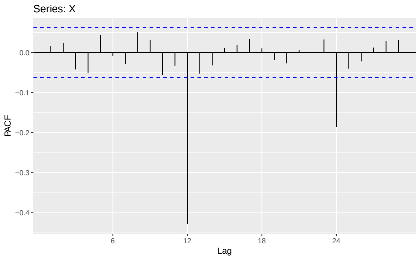


FIG.: PACF

Pure seasonal ARMA model

Pure seasonal ARMA$(P, Q)_s$

A pure seasonal **ARMA**$(P, Q)_s$ process X is a stationary process that is defined through

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)\omega_t$$

where $\omega \sim WN(0, \sigma^2)$, Φ_P is a polynomial of order P, Θ_Q is a polynomial of order Q and Φ_P and Θ_Q have no common factors.

Exercise

1. Verify that the pure seasonal ($s = 12$) ARMA$(0\,,1)_{12}$ (this is an MA$(1)_{12}$) has a ACF given by

$$\rho(12) = \theta/(1+\theta^2)$$
$$\rho(h) = 0 \text{ otherwise.}$$

2. Verify that the pure seasonal ($s = 12$) ARMA$(1\,,0)_{12}$ (this is an AR$(1)_{12}$) has a ACF given by

$$\rho(12k) = \phi^k \quad \text{for } k = 1, \ldots$$
$$\rho(h) = 0 \text{ otherwise.}$$

ARMA$(p, q) \times (P, Q)_s$

In general, we will mix seasonal and non-seasonal operators to build **multiplicative seasonal ARMA** : ARMA$(p, q) \times (P, Q)_s$.

ARMA$(0, 1) \times (1, 0)_{12}$

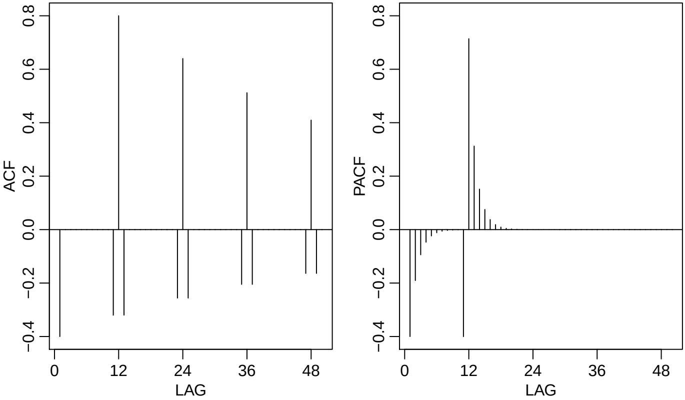


FIG.: ACF and PACF of the process $(1 - 0.8B^{12})X_t = (1 - 0.5B)\omega_t$ see [SS10]

SARIMA

Multiplicative seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$

A multiplicative seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$ process X is a process that is defined through

$$\Phi_P(B^s)\Phi(B)(1 - B^s)^D(1 - B)^d X_t = \Theta_Q(B^s)\Theta(B)\omega_t$$

where

- $\omega \sim WN(0, \sigma^2)$,
- Φ_P is a polynomial of order P
- Θ_Q is a polynomial of order Q
- Φ is a polynomial of order p
- Θ is a polynomial of order q

Model building

To choose p,q,P,Q,d,D

- First difference sufficiently to get to stationarity.
- Then find suitable orders for ARMA or seasonal ARMA models for the differenced time series. The ACF and PACF is again a useful tool here.
- Select few models, compare their AICc and BIC
- Finally conduct a diagnosis check for the residuals of the select model.

References I

Peter J Brockwell and Richard A Davis, *Time series : theory and methods*, Springer Science & Business Media, 2013.

Robert B Cleveland, William S Cleveland, and Irma Terpenning, *Stl : A seasonal-trend decomposition procedure based on loess*, Journal of Official Statistics **6** (1990), no. 1, 3.

Spyros Makridakis, Steven C Wheelwright, and Rob J Hyndman, *Forecasting methods and applications*, John Wiley & Sons, 2008.

Robert H Shumway and David S Stoffer, *Time series analysis and its applications : with r examples*, Springer Science & Business Media, 2010.