Introduction to time series analysis and forecasting

Hugues GARNIER

hugues.garnier@univ-lorraine.fr

Course outline

Introduction to time series analysis and forecasting

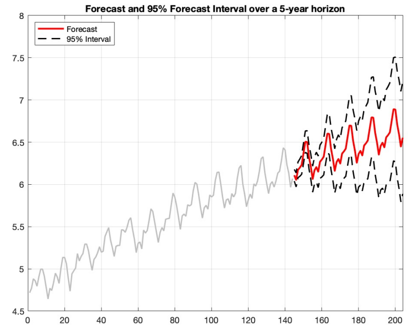
I. Main characteristics of time series data

II. Time series decomposition

III. Basic time series modelling and forecasting methods

IV. Stochastic time series modelling and forecasting:
The Box-Jenkins method for ARIMA models

Classification of time series forecasting methods



Forecasting methods

- Decomposition methods
  - Regression methods
- Smoothing methods
  - Moving average
  - Exp. smoothing
- The Box-Jenkins method
  - ARIMA models
  - SARIMA models
- Deep learning methods
  - LSTM models
  - GRU models, etc

- ARIMA methodology of forecasting is different from most methods because it does not assume any particular patterns in the historical data of the time series to be forecast

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

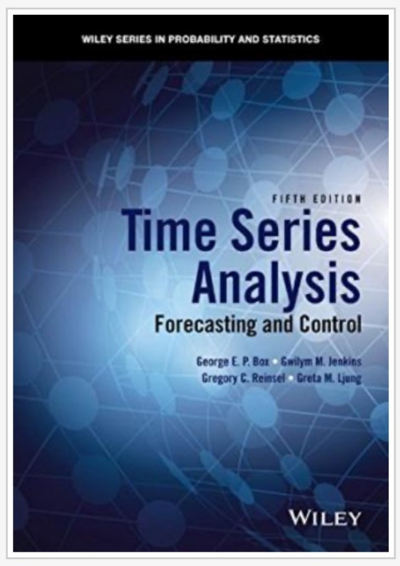
The Box-Jenkins method for ARIMA models

- It uses an interactive approach of identifying a possible model from a general class of models, named ARIMA

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d(y_t)=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$$

- The chosen model is then checked against the historical data to see if it accurately describes the series

- The Box-Jenkins method
  - has been remarkably successful
  - has excellent performance on ***small data*** sets
  - remains quite close to the performance of recent cutting edge methods



ARIMA models

- **A**uto**R**egressive **I**ntegrated **M**oving **A**verage (ARIMA) models were popularized by George Box and Gwilym Jenkins in the early 1970s

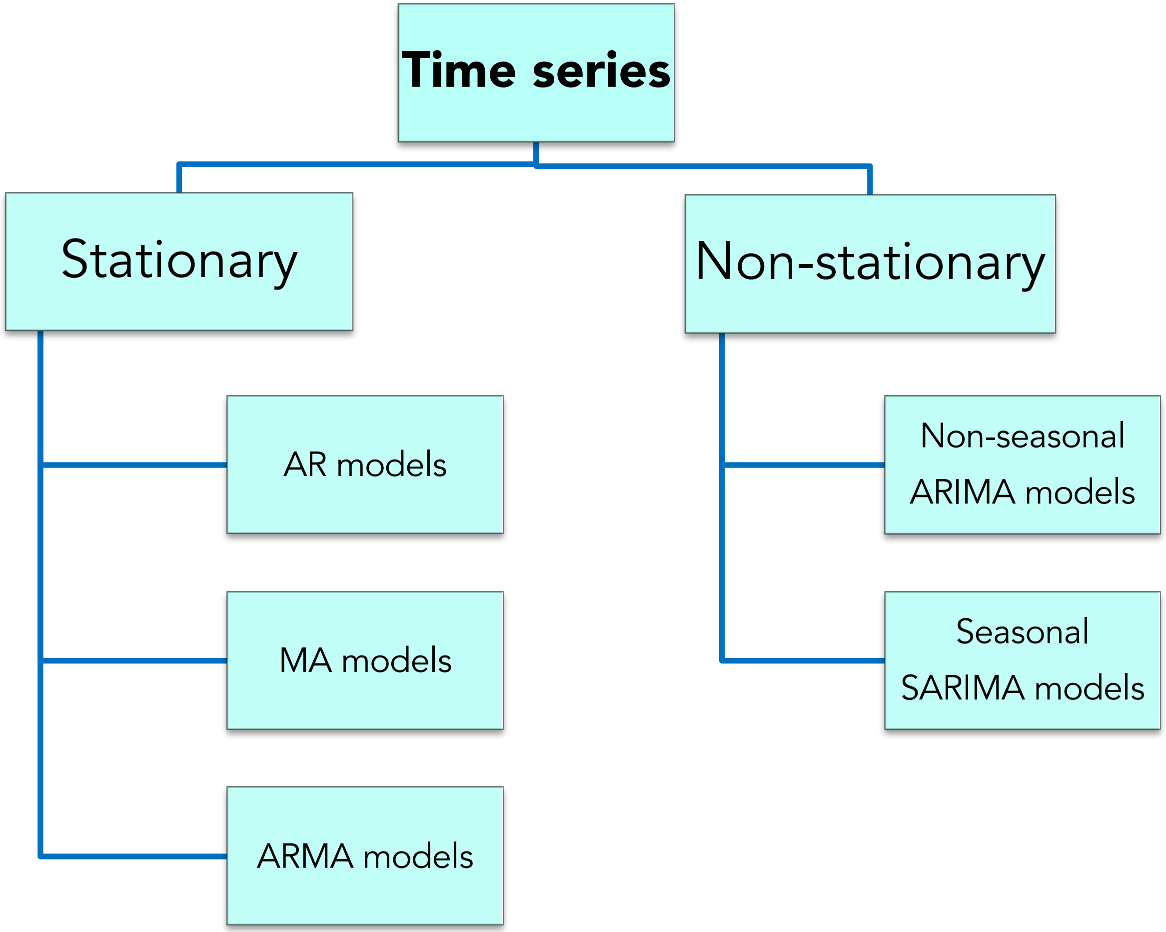
George E. P. Box   Gwilym M. Jenkins

- ARIMA models rely heavily on autocorrelation patterns in the data

- ARIMA models do not involve independent variables in their construction
  - They make use of the information in the series itself to generate forecasts

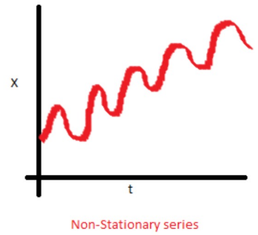
Family of ARIMA models

- ARIMA models are a class of ***black-box*** models that is capable of representing stationary ***as well as*** non-stationary time series

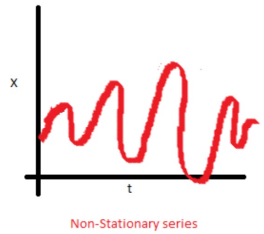


Major assumption: **stationarity** of the time series

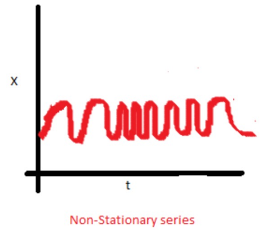
- The properties of one section of a data are much like the properties of the other sections. The future is "similar" to the past *(in a probabilistic sense)*
- A stationary time series has
  - no trend / no seasonality



  - no systematic change in variation



  - no periodic fluctuations



- One of the first steps in the Box-Jenkins method is to transform a non-stationary time series into a stationary one *(by using a detrending or differencing method)*

Key statistics for time series analysis:
Autocorrelation and partial autocorrelation functions

- *Autocorrelation* and *partial autocorrelation plots* are heavily used in time series analysis and forecasting
- These are plots that graphically summarize the strength of a relationship with an observation in a time series with observations at prior time instants
- The difference between autocorrelation and partial autocorrelation plots can be difficult and confusing for beginners to time series forecasting
- Plots of the autocorrelation and partial autocorrelation functions for a time series ***tell a very different story*** and are very useful to select the order of an ARIMA model

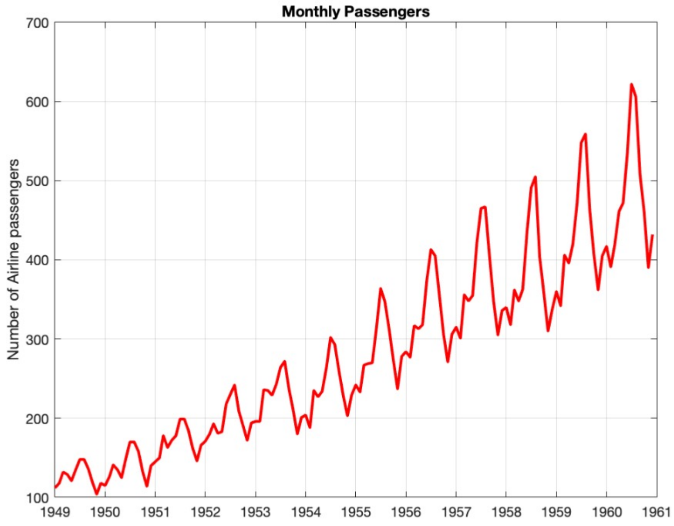
Autocorrelation function (ACF)

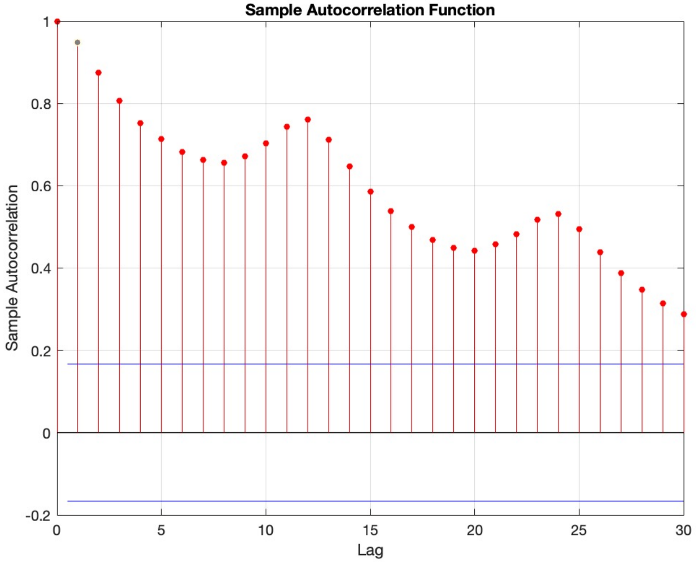
- Statistical correlation summarizes the strength of the relationship between two different variables

- We can calculate the correlation for time series observations with observations with previous time instants, called lags. This is called an autocorrelation

- A plot of the autocorrelation of a time series in terms of lags is called the AutoCorrelation Function, or its acronym ACF

- Sample ACF at lag h, *denoted as* $\gamma_y(h)$, measures the linear correlation between y_t and y_{t+h}





ACF: stationarity case

- Autocovariance function of a stationary time series $\{y_t\}$

$$\gamma_y(h) = \text{Cov}(y_{t+h}, y_t) = \text{E}[(y_{t+h} - \mu)(y_t - \mu)] \quad |h| < N$$

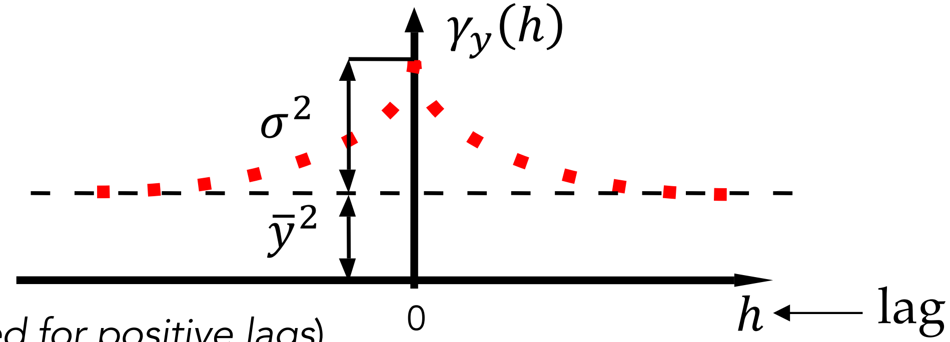
with the following 3 properties

1. $\gamma_y(0) \geq 0,$

2. $|\gamma_y(h)| \leq \gamma_y(0)$

3. $\gamma_y(h) = \gamma_y(-h)$

⇒ even function. ACF is usually plotted for positive lags)



- Autocorrelation function of a stationary time series $\{y_t\}$

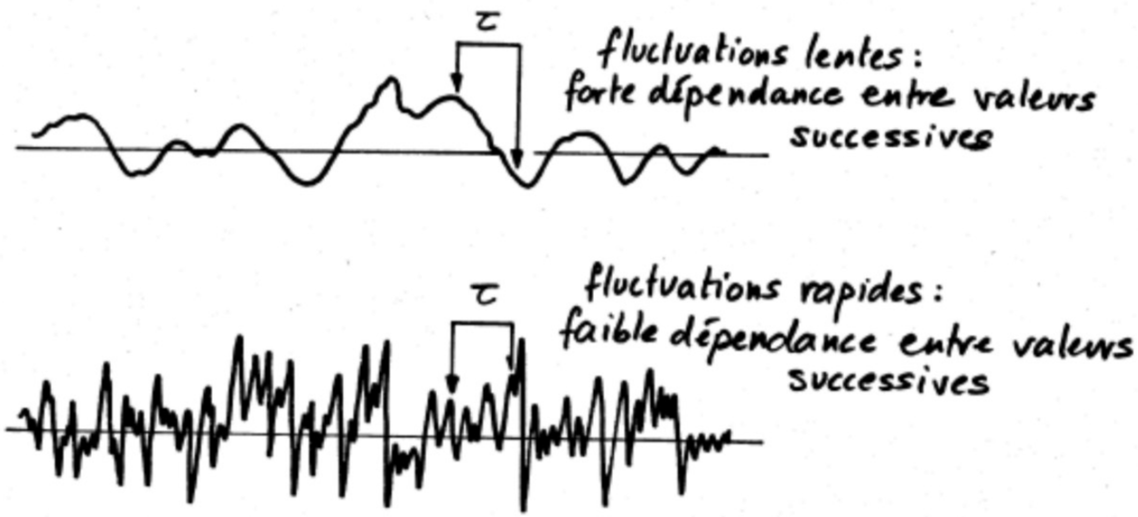
$$\rho_y(h) = \frac{\gamma_y(h)}{\gamma_y(0)} \qquad 0 \leq h < N$$

with all the properties of the autocovariance function, except $\rho_y(0) = 1$

- It measures the linear correlation between y_t and y_{t+h}

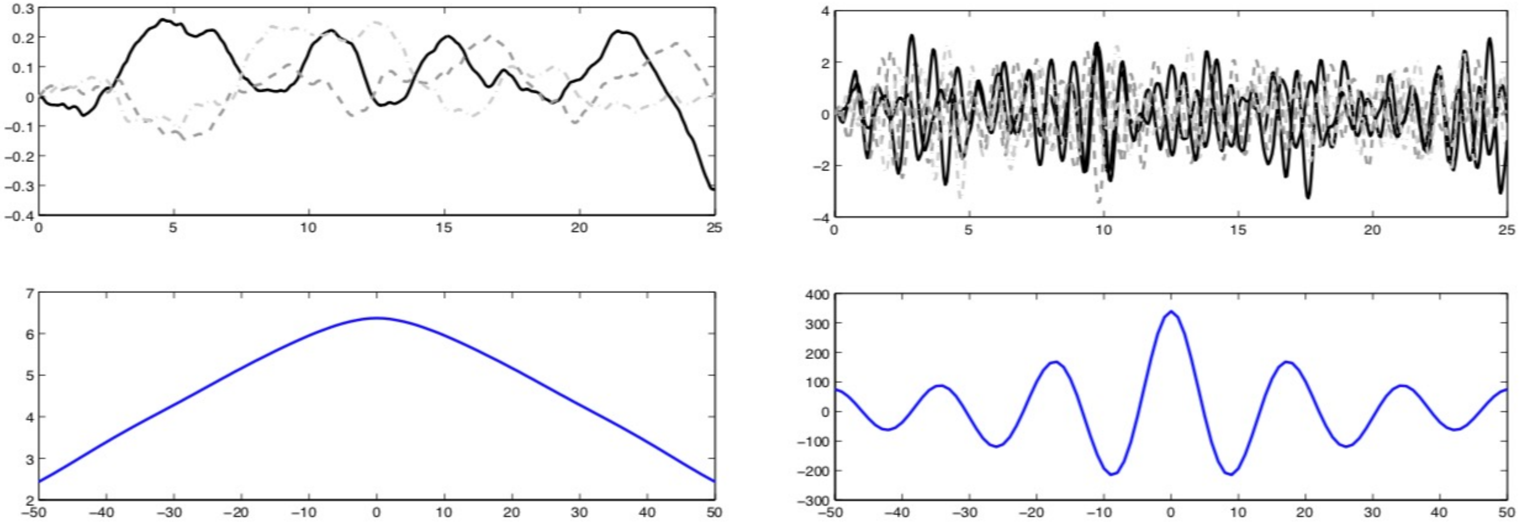
Autocorrelation function (ACF)

- ACF: *measures the speed of variation of temporal evolutions*
  - we compare the time series with itself but shifted by τ (*or h*)
  - it allows us to see how the time series at a given time is influenced (*linear autocorrelation*) by what happened at a previous time



Autocorrelation function (ACF)

Slowly varying autocorrelation function – slowly varying process
Quickly varying autocorrelation function – quickly varying process



Sample statistics

- Given $\{y_1, \ldots, y_N\}$ observations of a stationary time series $\{y_t\}$, estimate the sample mean, variance, autocovariance and ACF
  - Sample mean

$$\hat{\mu} = \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

  - Sample variance

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{\mu})^2$$

  - Sample autocovariance function

$$\hat{\gamma}_y(h) = \frac{1}{N}\sum_{j=1}^{N-h}\left(y_{j+h} - \bar{y}\right)\left(y_j - \bar{y}\right), \qquad 0 \leq h < N,$$

with $\hat{\gamma}_y(h) = \hat{\gamma}_y(-h), \qquad -N < h \leq 0$

  - Sample autocorrelation function (ACF)

$$\hat{\rho}_y(h) = \frac{\hat{\gamma}_y(h)}{\hat{\gamma}_y(0)}, \qquad |h| < N$$

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Sample ACF - Example

$y = [0\ 1\ 1\ 1\ 0]$ $\quad N=5$

$$\bar{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = 0.6$$

$$\hat{\gamma}_y(h) = \frac{1}{5}\sum_{j=1}^{5-h}(y_{j+h} - \bar{y})(y_j - \bar{y}), \qquad h = 0,1,2,3,4$$

$$\hat{\rho}_y(h) = \frac{\hat{\gamma}_y(h)}{\hat{\gamma}_y(0)}, \qquad h = 0,1,2,3,4$$

$$\hat{\rho}_y = [1\ -0.13\ -0.26\ -0.4\ \ 0.3]$$

In Matlab :

```
y=[0 1 1 1 0];
[rho_hat_y,Lag]=xcov(y,'norm');
stem(Lag,rho_hat_y)
Or
autocorr(y)
```

(Matlab econometrics toolbox)

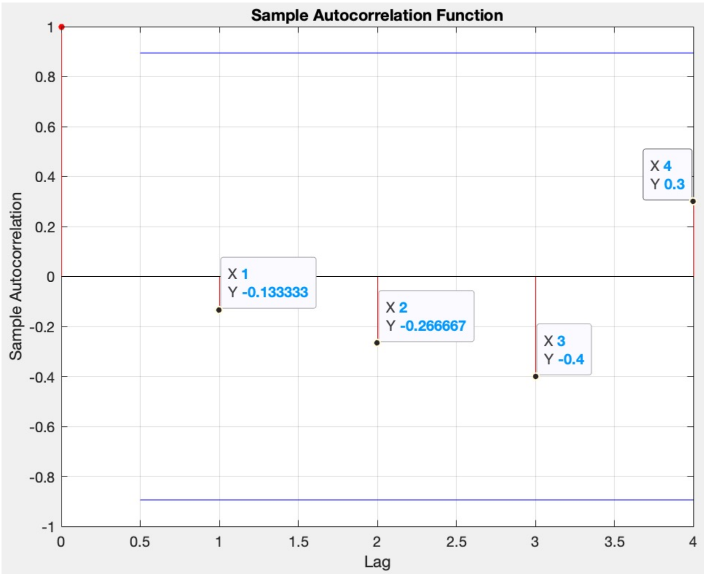
$$\hat{\gamma}_y(0) = \frac{1}{5}\sum_{j=1}^{5}(y_j - \bar{y})(y_j - \bar{y}) = 0.24$$

$$\hat{\gamma}_y(1) = \frac{1}{5}\sum_{j=1}^{4}(y_{j+1} - \bar{y})(y_j - \bar{y}) = -0.0320$$

$$\hat{\gamma}_y(2) = \frac{1}{5}\sum_{j=1}^{3}(y_{j+2} - \bar{y})(y_j - \bar{y}) = -0.0620$$

$$\hat{\gamma}_y(3) = \frac{1}{5}\sum_{j=1}^{2}(y_{j+3} - \bar{y})(y_j - \bar{y}) = -0.0960$$

$$\hat{\gamma}_y(4) = \frac{1}{5}\sum_{j=1}^{1}(y_{j+4} - \bar{y})(y_j - \bar{y}) = 0.0720$$



Partial autocorrelation function (PACF)

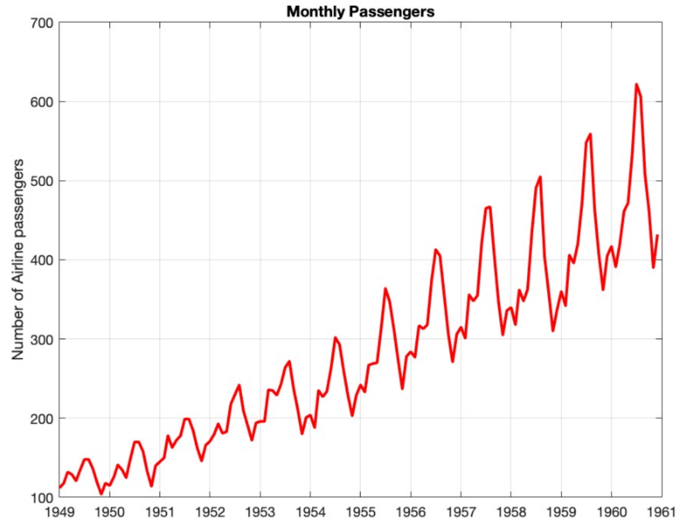
- The autocorrelation for an observation y_t and an observation at a prior time-instant y_{t+h} is comprised of both the direct correlation and indirect correlations between y_t and $y_{t+1}, y_{t+2}, \dots, y_{t+h-1}$

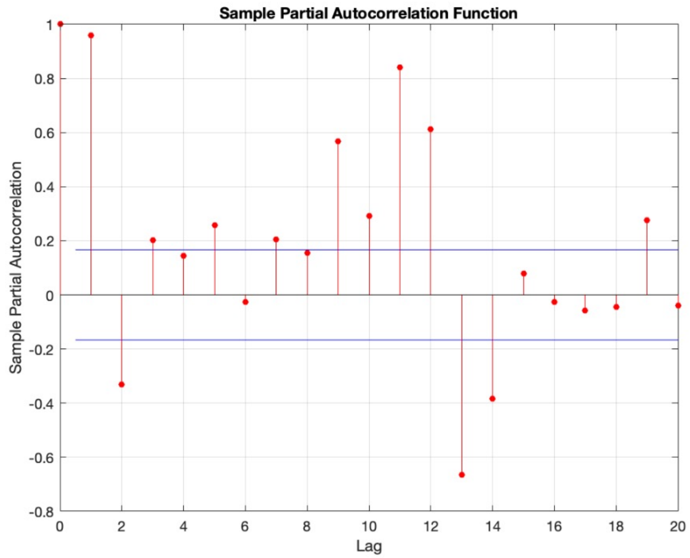
- These indirect correlations are a linear function of the correlation of the observation, with observations at intermediate time-instants

- It is these indirect correlations that the partial autocorrelation function tries to remove

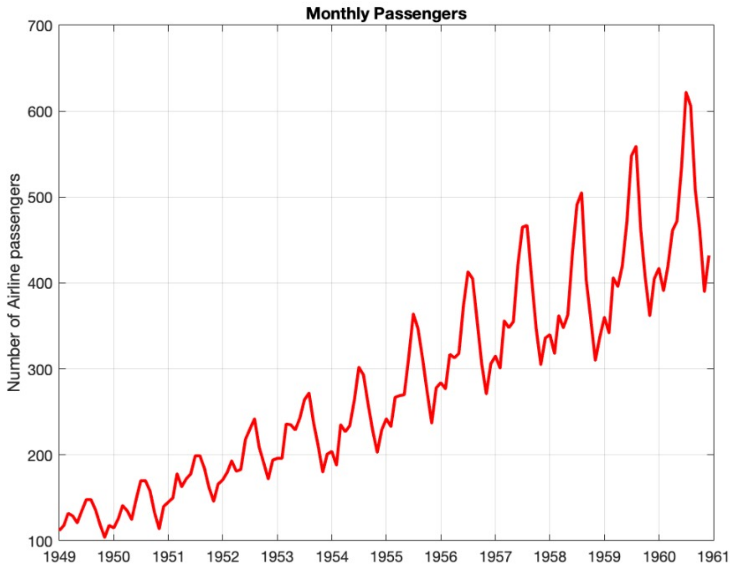
- A plot of the partial autocorrelation of a time series in terms of lags is called the Partial Autocorrelation Function, or by its acronym PACF

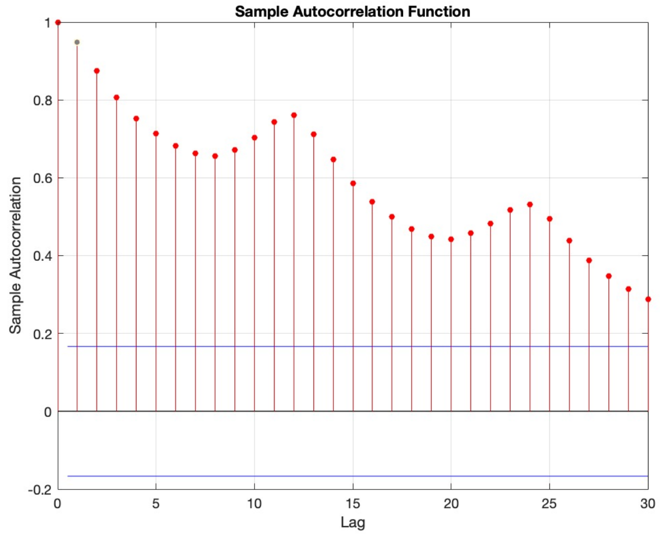
- Sample PACF at lag h, *denoted as* $\alpha_y(h)$, measures the linear correlation between y_t and y_{t+h}, but after statistically removing the effect of $y_{t+1}, y_{t+2}, \dots, y_{t+h-1}$

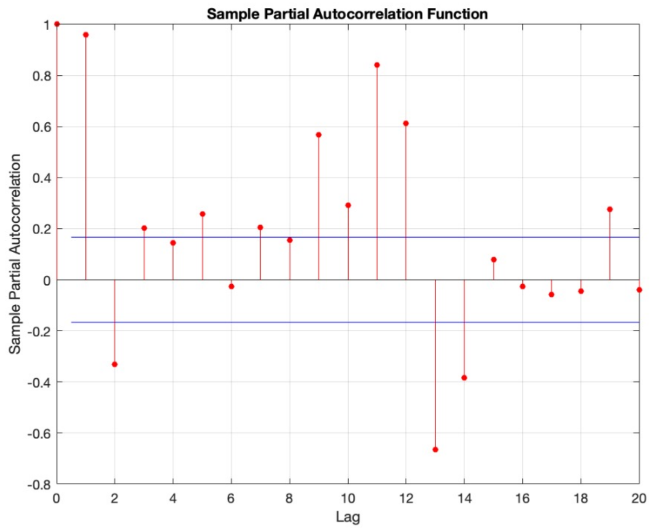




Plots of the ACF and PACF for a time series *tell a very different story - Example*



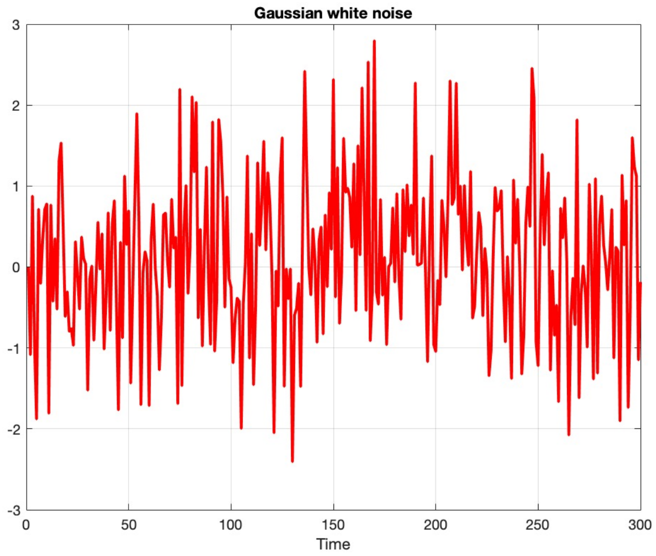


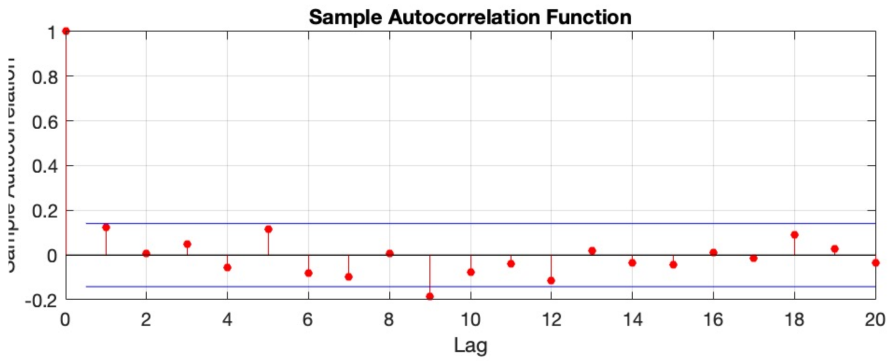


The white noise process

The most fundamental example of stationary process

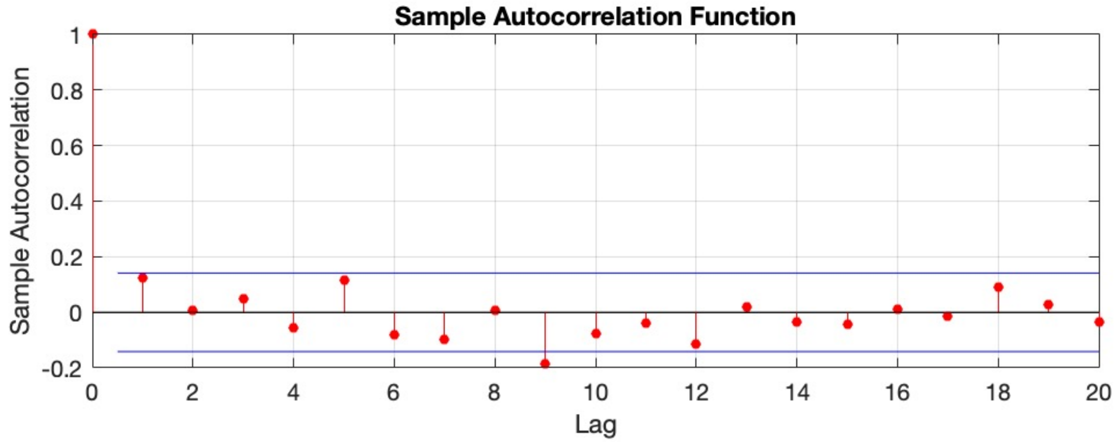
- A white noise process is a sequence of independent and identically distributed *(i.i.d)* random variables
  - The sequences are *uncorrelated*, have zero mean, and constant variance
  - A Gaussian white noise are *i.i.d* observations from $\mathcal{N}(0, \sigma^2)$
  - Because independence implies that its variables are uncorrelated at different times, its ACF looks like a Kronecker impulse





Sampling distribution of sample ACF

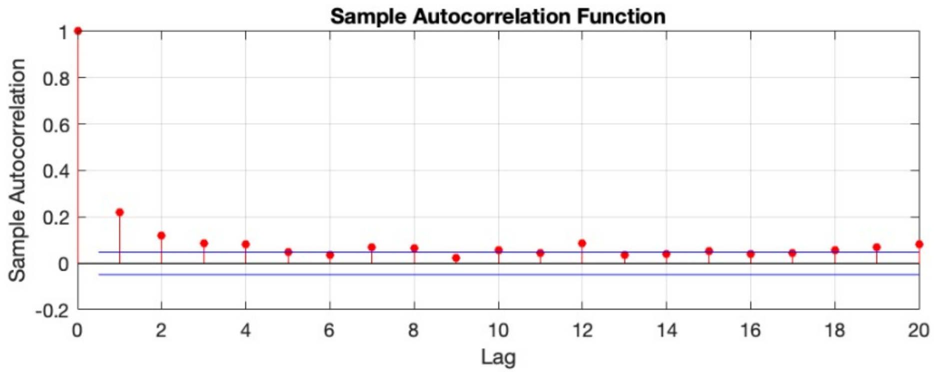
- Sampling distribution of ACF for a white noise is asymptotically Gaussian $\mathcal{N}\left(0, \frac{1}{N}\right)$
  - 95% of all ACF coefficients for a white noise must lie within $\pm \frac{1.96}{\sqrt{N}}$
  - It is common to plot horizontal limit lines at $\pm \frac{1.96}{\sqrt{N}}$ when plotting the ACF
- If $N = 125$, critical values at $\pm \frac{1.96}{\sqrt{125}} = \pm 0.175$
  - All ACF coefficients lie within these limits, confirming that the data are white noise (*more precisely, the data cannot be distinguished from white noise*)



H. Garnier

Properties of white noise process

- Best forecast of a white noise
  - If a time series is white noise, it is unpredictable and so there is nothing to forecast. Or more precisely, *the best forecast is its mean value which is zero*
- Whitening test of the residuals
  - At the validation stage of the Box-Jenkins methodology, we will check whether the forecast errors (=*the residuals*) are a white noise by plotting its sample ACF



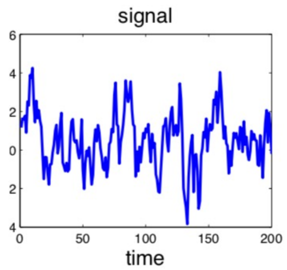
Sample ACF shows some significant autocorrelations at lags 1, 2, 3 and 4. This shows the residuals are not white here

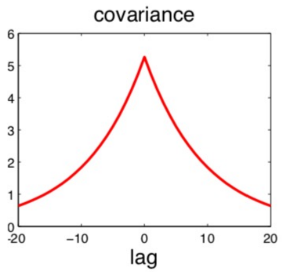
  - If the residual ACF does not resemble to the ACF of a white noise, it suggests that improvements could be made to the predictive model
  - If the residual ACF resembles to the ACF of a white noise, the modelling procedure is finished. There is nothing else to capture in the residuals and the estimated ARIMA model can be used for forecast

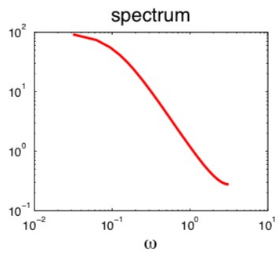
Models for stationary random signals or time series

Signaux aléatoires stationnaires

- Modèles non paramétriques
  - Evolution temporelle
  - Fonction d'autocorrélation
  - Densité spectrale de puissance
- Modèles paramétriques
  - Modèles AR, MA, ARMA







$$y_t = \frac{\Theta(L)}{\Phi(L)} \varepsilon_t$$

General linear parametric model of stationary time series

- Box and Jenkins in 1970 *(following Yule and Slutsky 1927)*
  - Many time series *(or their derivatives)* can be considered as a special class of stochastic processes: *(weakly)* stationary stochastic processes
    - First two moments are finite and constant over time
    - Defined completely by the mean, variance and autocorrelation function

- General parametric model of stationary stochastic processes *(Wold 1938)*
  - All *(weakly)* stationary stochastic processes can be written as

$$y_t = c + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i} + \varepsilon_t$$

where c is a constant and ε_t is a white Gaussian noise

  - ε_t is often called the *innovation process* because it captures all new information in the series at time *t*

Lag or backward shift L operator

- The ***Lag or backward shift*** operator, L, is defined as

$$L\,\varepsilon_t = \varepsilon_{t-1}$$

$$L^i\,\varepsilon_t = \varepsilon_{t-i}$$

- The general linear model of a stationary stochastic process can be written as

$$y_t = c + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i} + \varepsilon_t$$

$$y_t = c + \psi(L)\,\varepsilon_t$$

$$\Psi(L) = 1 + \sum_{i=1}^{+\infty} \psi_i L^i$$

- This model has an infinite-degree polynomial $\Psi(L)$ with infinite coefficients which cannot be estimated from a finite amount of data in the time series 😩

Towards AR, MA and ARMA models for stationary time series

- If $\Psi(L)$ is a rational polynomial, we can write it (at least approximately) as the quotient of two finite-degree polynomials

$$\Psi(L) = \frac{\Theta(L)}{\Phi(L)}$$

$$\Theta(L) = 1 + \theta_1 L^1 + \cdots + \theta_q L^q$$
$$\Phi(L) = 1 - \phi_1 L^1 - \cdots - \phi_p L^p$$

(Matlab Econometrics toolbox notations)

- ***Wold's theorem***: every stationary stochastic process can be written as

$$y_t = c + \frac{\Theta(L)}{\Phi(L)} \varepsilon_t$$

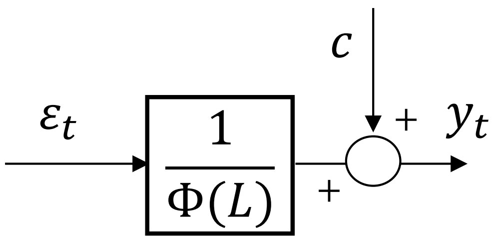
  - which has a finite number $(p+q)$ of coefficients

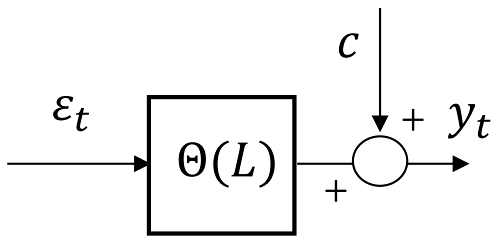
- This leads to the use of parsimonious models : **AR, MA** and **ARMA** models
  - They are most useful for practical applications since these models can be quite easily estimated from a finite amount of data in the time series

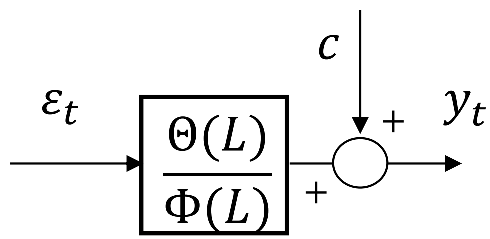
Family of ARMA models for stationary time series

- ARMA models: a way to "see" stationary time series as *filtered white noise*
  - The filter takes different forms according to the time series properties

AR models	MA models	ARMA models







$$y_t = c + \frac{1}{\Phi(L)}\varepsilon_t$$

$$y_t = c + \Theta(L)\varepsilon_t$$

$$y_t = c + \frac{\Theta(L)}{\Phi(L)}\varepsilon_t$$

c is a constant *(mean of the time series)*

$\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$

AutoRegressive models: AR(p) models

- An autoregressive model of order *p*, AR(p), is defined by *(Yule 1927)*

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t$$

  – where $p \geq 1$, c is a constant and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$

- It can also be written in Lag-operator polynomial form:

$$\Phi(L)\,(y_t - c) = \varepsilon_t$$

$$\Phi(L) = 1 - \phi_1 L^1 - \cdots - \phi_p L^p$$

(Matlab Econometrics toolbox notations)

- Stationarity conditions
  – An AR(p) process is stationary if all roots of $\Phi(L)$ are outside the unit circle
- Special case
  – if one or more roots lie on the unit circle (i.e., have absolute value of one), the model is called a ***unit root process*** model, which is non-stationary
    - When *p*=1 and $c = 0, \phi_1$=1, $y_t = y_{t-1} + \varepsilon_t$ is a non-stationary ***random walk***, which is a unit root process

Different forms of an AR(2) model
Example

- An autoregressive model AR(2) of order 2 is given

$$y_t = -0.5y_{t-1} - 0.9y_{t-2} + \varepsilon_t$$

- It can also be written in Lag-operator polynomial form:

$$y_t + 0.5y_{t-1} + 0.9y_{t-2} = \varepsilon_t$$

$$(1 + 0.5L^1 + 0.9L^2)y_t = \varepsilon_t$$

$$\Phi(L)y_t = \varepsilon_t$$

$$\Phi(L) = 1 + 0.5L^1 + 0.9L^2$$

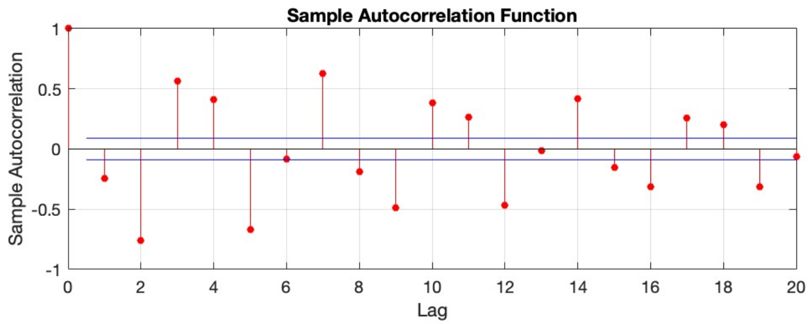
- Stationarity condition
  - This AR(2) process is stationary since the two roots of $\Phi(L)$ are outside the unit circle ($L_{1,2} = -0.28 \pm 1.02i$)
  - *Note that the polynomial $\Phi(L)$ is written in a different form than usually used in Control Engineering or Signal Processing where the backward operator q^{-1} is used so that the polynomial would be $\Phi(q^{-1}) = 1 + 0.5q^{-1} + 0.9q^{-2}$. With this negative power notation for the polynomial, the filter would be stable if the roots of $\Phi(q^{-1})$ are inside the unit circle. Do not be confused by the Lag-operator polynomial form used here and apply the appropriate rule to test the stationarity of the AR process!*

Properties of AR(p) process

- Autocorrelation function

$$\lim_{h \to +\infty} \gamma_y(h) = 0$$

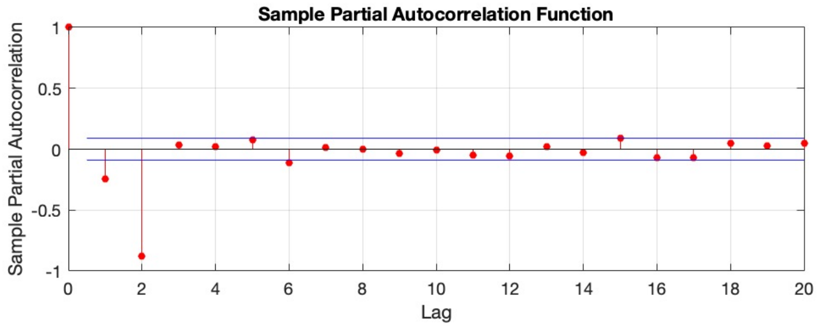
  - The sample ACF exponentially decreases to 0 when $h \to +\infty$



- Partial autocorrelation function

$$\alpha_y(h) = \phi_h \text{ for } |h| = p$$
$$\alpha_y(h) = 0 \text{ for } |h| > p$$

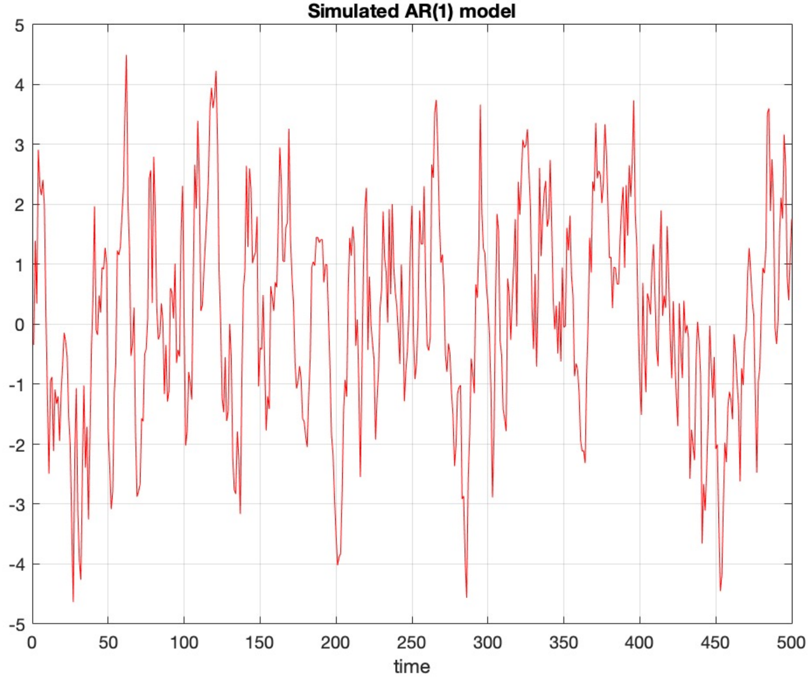
  - The sample PACF of an AR(p) process cuts off after p lags

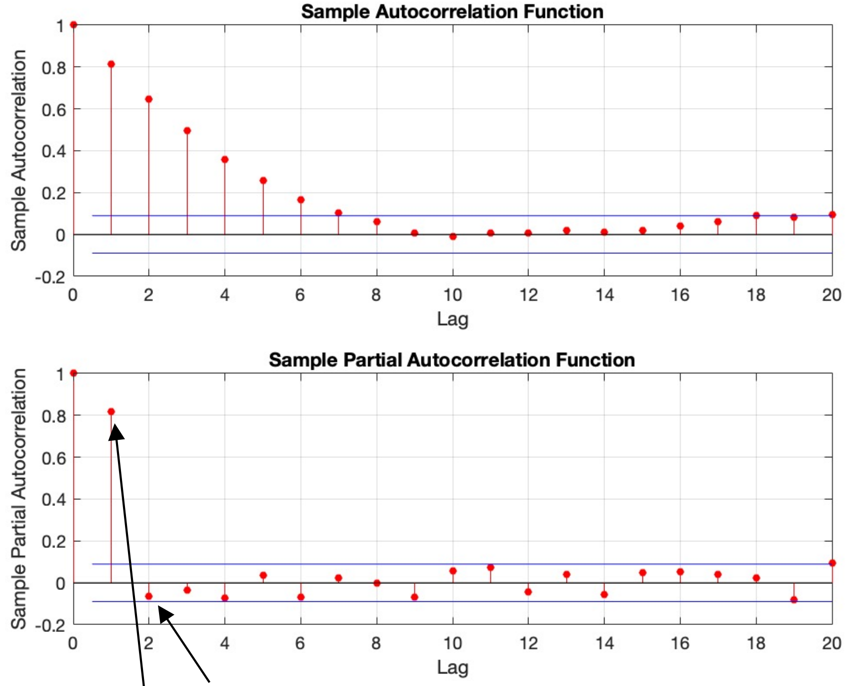


- Order selection of an AR process
  - *PACF is the plot to be used to select the order p of an AR process*

AR(1) process example:

$$y_t = 0.8y_{t-1} + \varepsilon_t$$



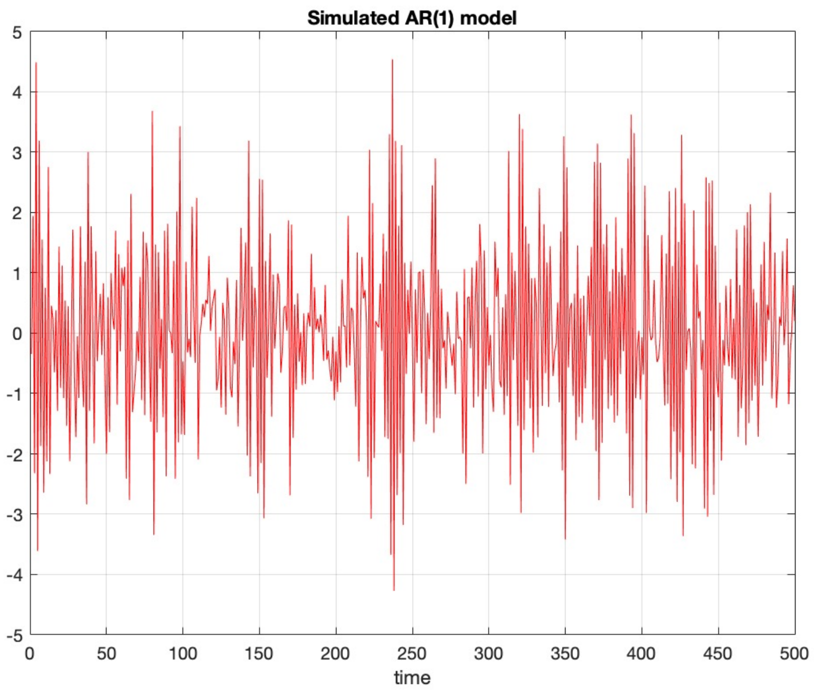


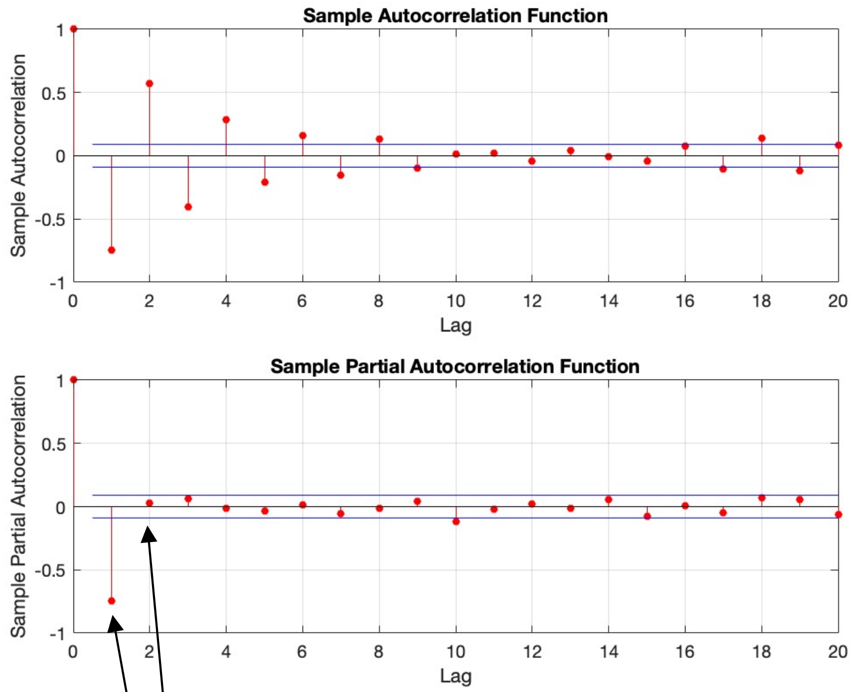
PACF cuts off after 1 lag ⇒ AR(1) process

PACF(1) = ϕ_1=0.8

AR(1) process example:

$$y_t = -0.8 y_{t-1} + \varepsilon_t$$



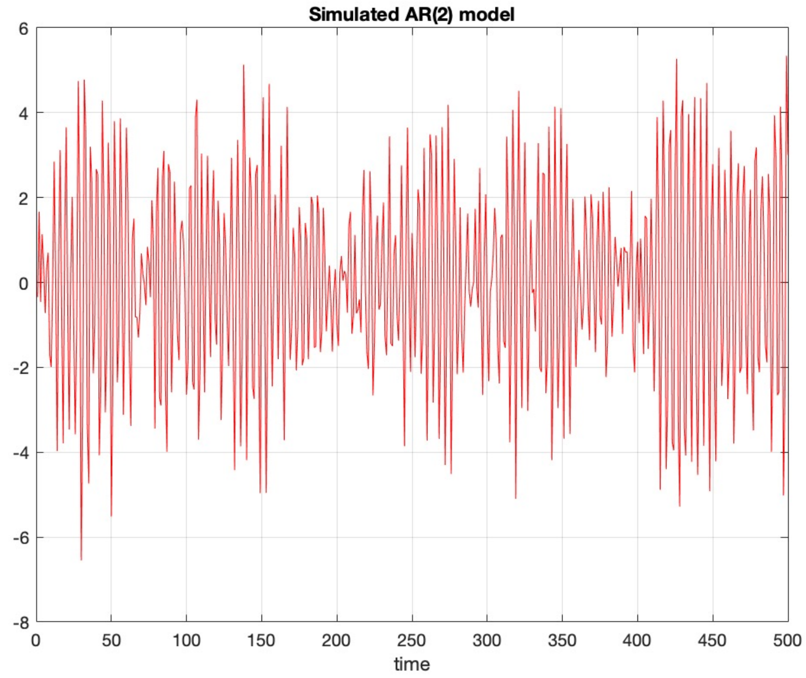


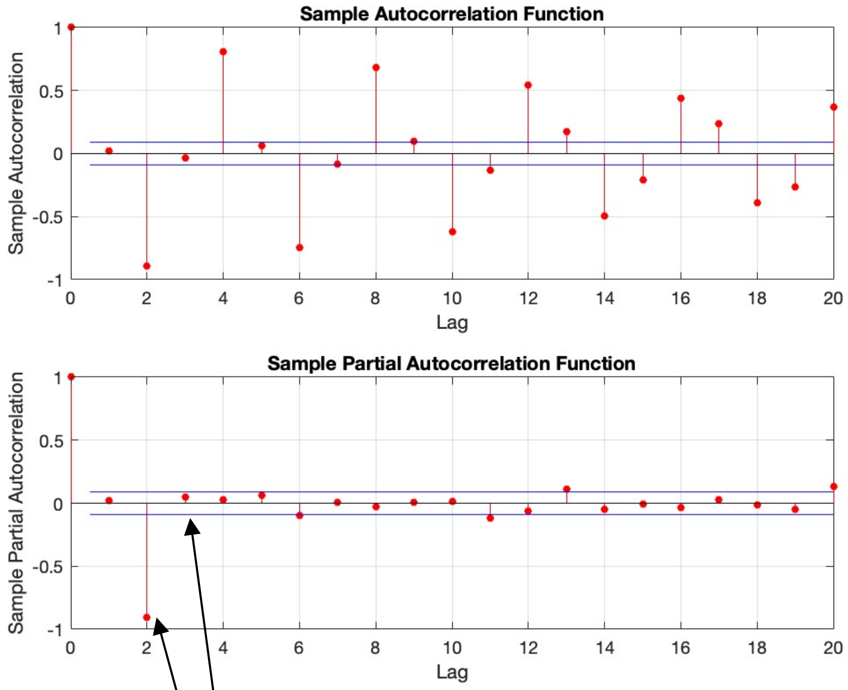
PACF cuts off after 1 lag ⇒ AR(1) process

PACF(1) = ϕ_1*=-0.8*

AR(2) process example:

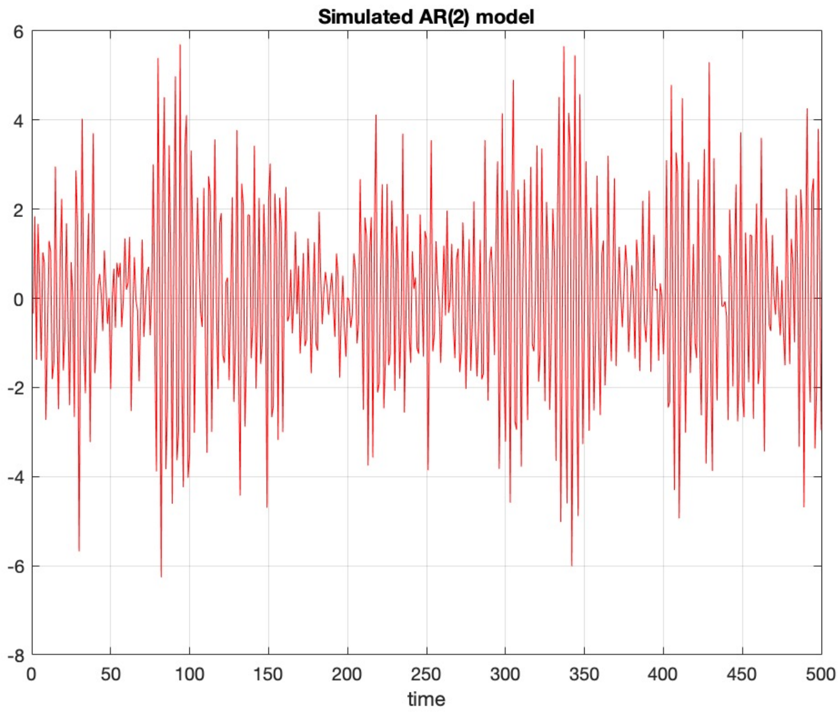
$$y_t = -0.9y_{t-2} + \varepsilon_t$$

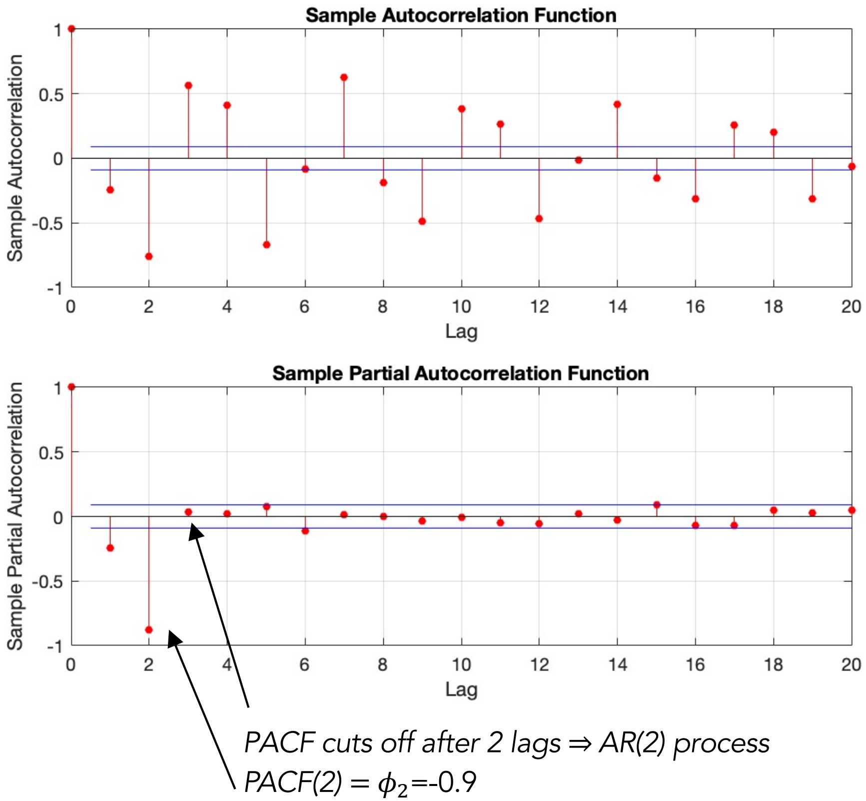




AR(2) process example:

$$y_t = -0.5y_{t-1} - 0.9y_{t-2} + \varepsilon_t$$





UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Moving Average models: MA(q) models

- A moving average model of order q , MA(q), is defined by (Slutsky 1927)

$$y_t = c + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

  – where $q \geq 1$, c is a constant and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$

- It can also be written in Lag polynomial form:

$$(y_t - c) = \Theta(L)\varepsilon_t$$

$$\Theta(L) = 1 + \theta_1 L^1 + \cdots + \theta_q L^q$$

(Matlab Econometrics toolbox notations)

- Stationarity and invertibility conditions
  – An MA(q) process is always stationary *(to the second order)*
  – An MA(q) process is invertible if all its roots are outside the unit circle *(required to be able to compute forecast)*

Moving Average models and related methods should not be confused with Moving Average smoothing methods !

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Different forms of an MA(2) model
Example

- A moving average model MA(2) of order 2 is given

$$y_t = \varepsilon_t - 0.8\varepsilon_{t-1} + 0.5\varepsilon_{t-2}$$

- It can also be written in Lag polynomial form:

$$y_t = (1 - 0.8L^1 + 0.5L^2)\varepsilon_t$$

$$y_t = \Theta(L)\varepsilon_t$$

$$\Theta(L) = 1 - 0.8L^1 + 0.5L^2$$

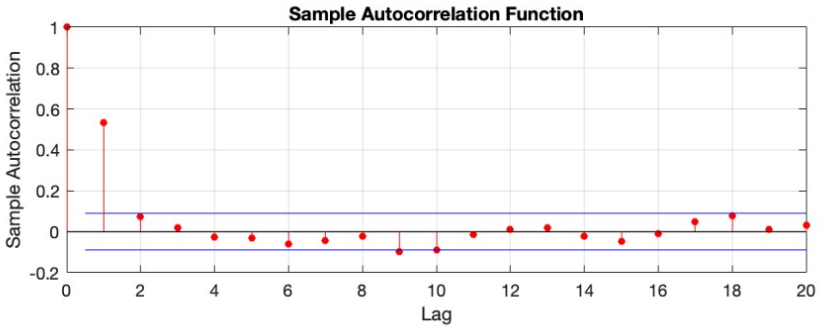
- Stationarity and invertibility conditions
  - The MA(2) process is stationary *(always)* and invertible since the two roots of $\Theta(L)$ are outside the unit circle ($L_{1,2} = 0.8 \pm 1.16i$)

Properties of MA(q) process

- Autocorrelation function

$$\gamma_y(h) = 0 \text{ for } |h| > q$$

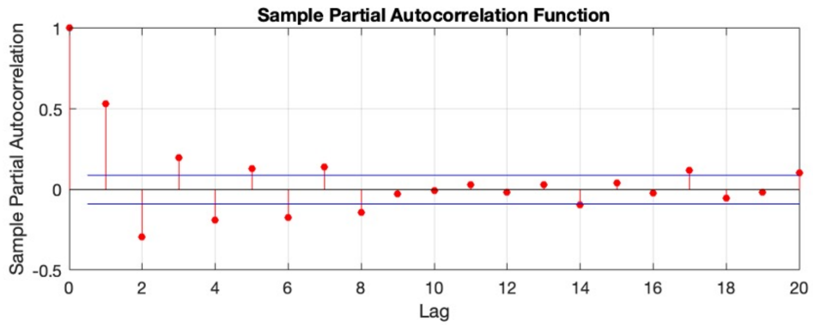
  - The sample ACF of an MA(q) process cuts off after q lags



- Partial autocorrelation function

$$\lim_{h\to+\infty} |\alpha_y(h)| = 0$$

  - The absolute value of the sample PACF exponentially decreases to *0* when $h \to +\infty$

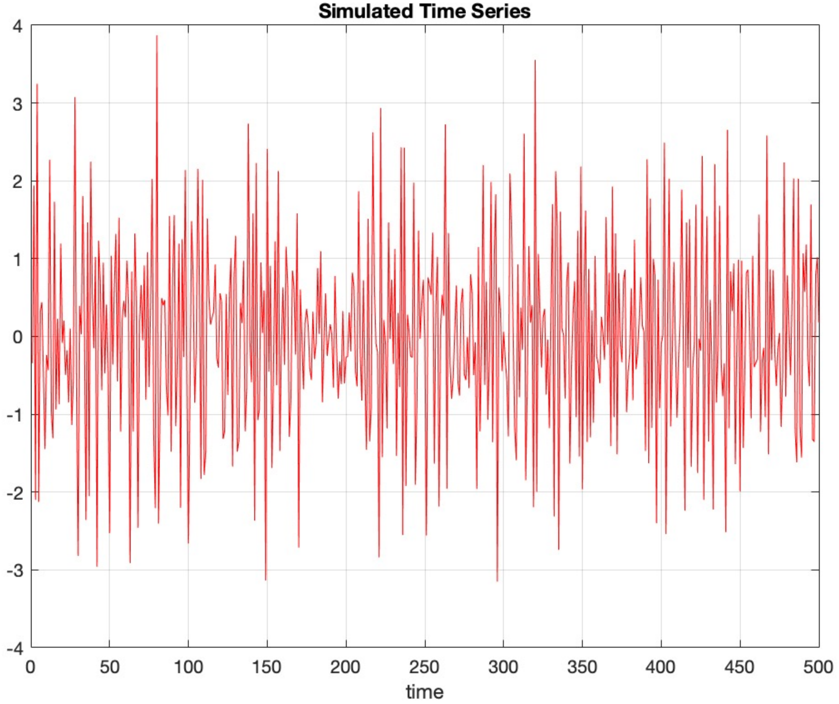


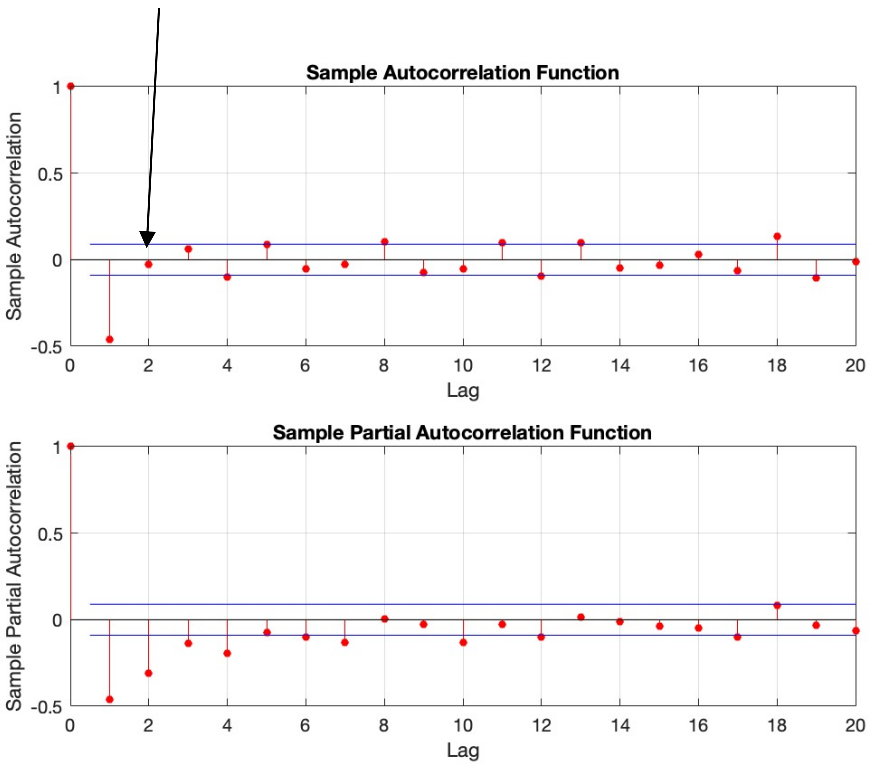
- Order selection of an MA(q) process
  - *ACF is the plot to be used to select the order q of an MA process*

MA(1) process example:

$$y_t = \varepsilon_t - 0.8\varepsilon_{t-1}$$

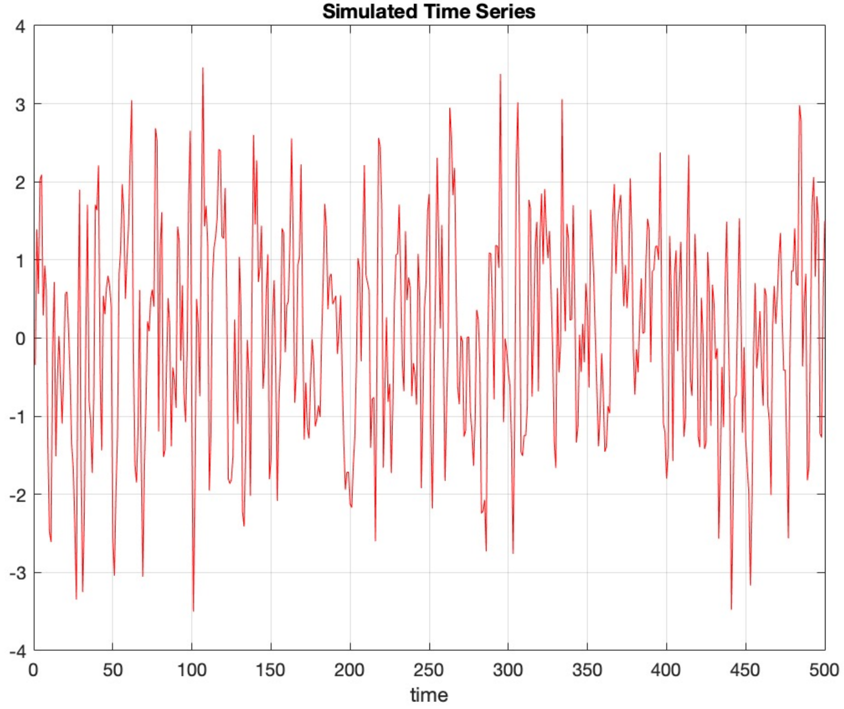
ACF cuts off after 1 lag ⇒ MA(1) process



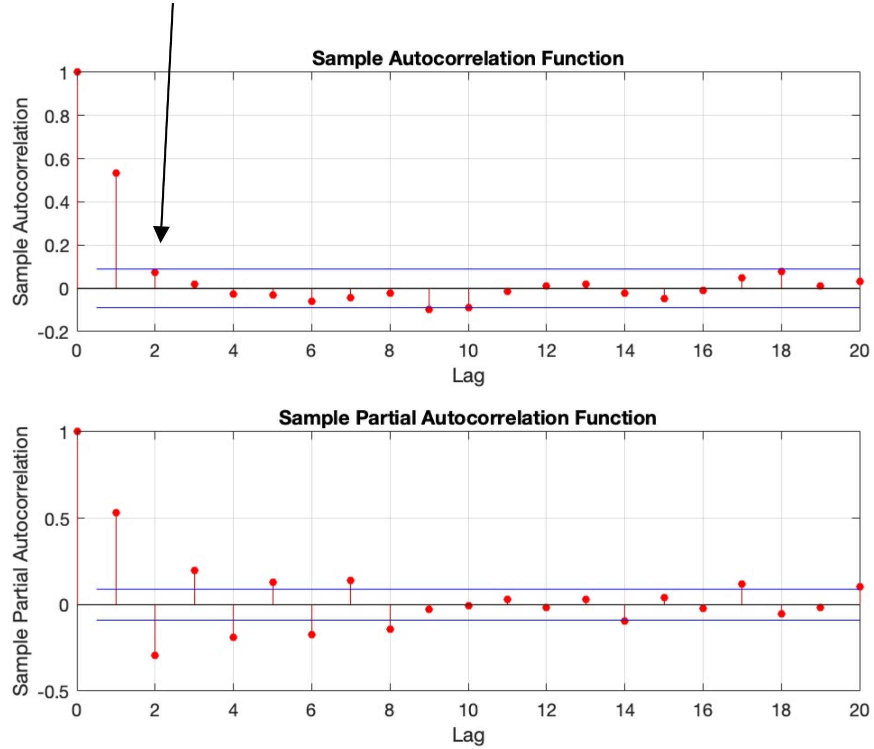


MA(1) process example:

$$y_t = \varepsilon_t + 0.8\varepsilon_{t-1}$$



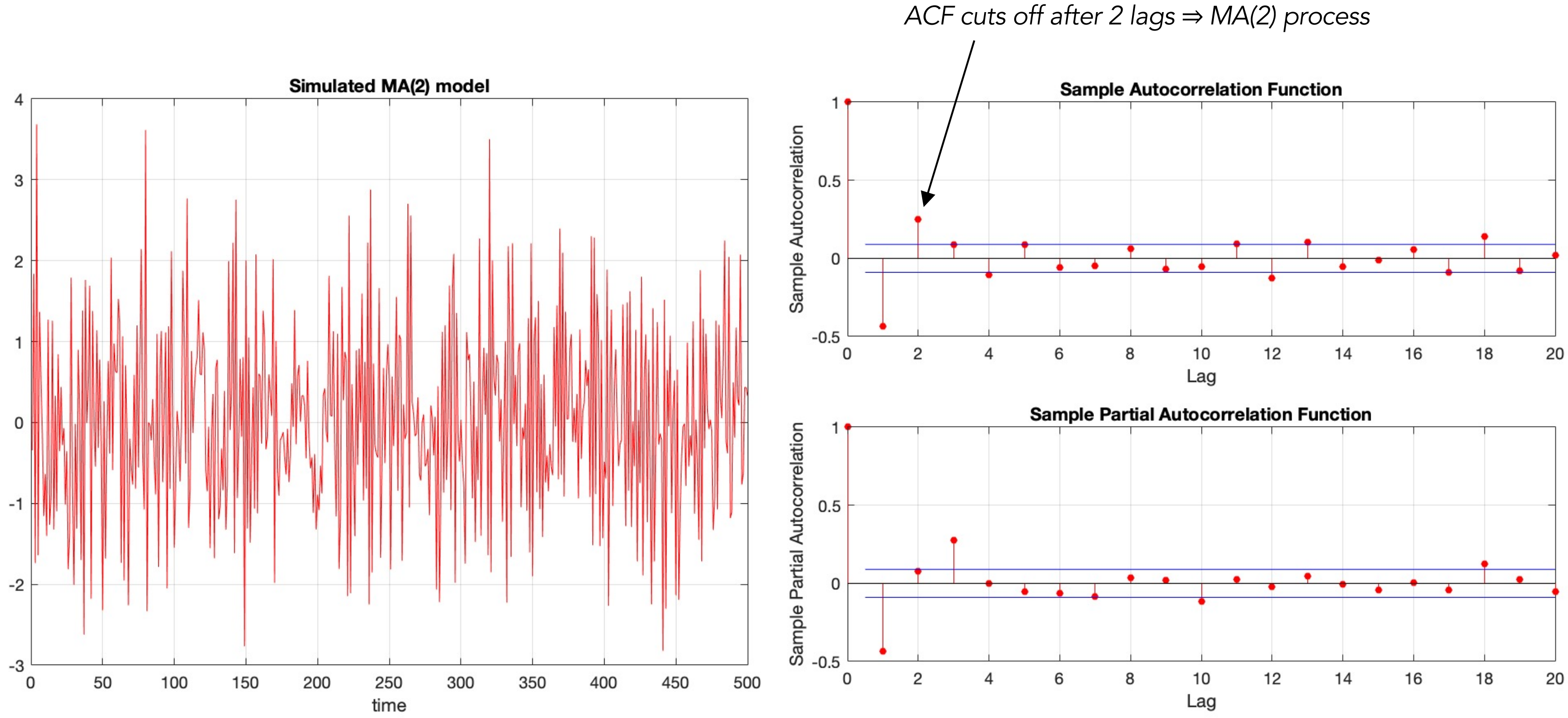
ACF cuts off after 1 lag ⇒ MA(1) process



H. Garnier

MA(2) process example:

$$y_t = \varepsilon_t - 0.5\varepsilon_{t-1} + 0.4\varepsilon_{t-2}$$



Moving Average Autoregressive models: ARMA(p,q) models

- An ARMA(p,q) of order p and q is defined by (Slutsky 1927)

$$y_t = \theta_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

  - where $p \geq 1$, $q \geq 1$, $\theta_0 = c\Phi(L)$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$

- It can also be written in Lag-operator polynomial form:

$$\Phi(L) y_t = \theta_0 + \Theta(L) \varepsilon_t$$

or

$$\boxed{\Phi(L)(y_t - c) = \Theta(L) \varepsilon_t}$$

(Matlab econometrics toolbox notations)

$$\Theta(L) = 1 + \theta_1 L^1 + \cdots + \theta_q L^q$$
$$\Phi(L) = 1 - \phi_1 L^1 - \cdots - \phi_p L^p$$

- Stationarity and invertibility conditions
  - An ARMA(p,q) process is stationary if all roots of $\Phi(L)$ are outside the unit circle
  - An ARMA(p,q) process is invertible if all roots of $\Theta(L)$ are outside the unit circle

Different forms of an ARMA(1,1) model
Example

- A moving average model ARMA(1,1) is given

$$y_t = 0.8y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$$

- It can also be written in Lag-operator polynomial form:

$$y_t - 0.8y_{t-1} = \varepsilon_t - 0.5\varepsilon_{t-1}$$
$$(1 - 0.8L)y_t = (1 - 0.5L)\varepsilon_t$$

$$\Phi(L)y_t = \Theta(L)\varepsilon_t$$
$$\Theta(L) = (1 - 0.5L)$$
$$\Phi(L) = (1 - 0.8L)$$

Properties of ARMA process

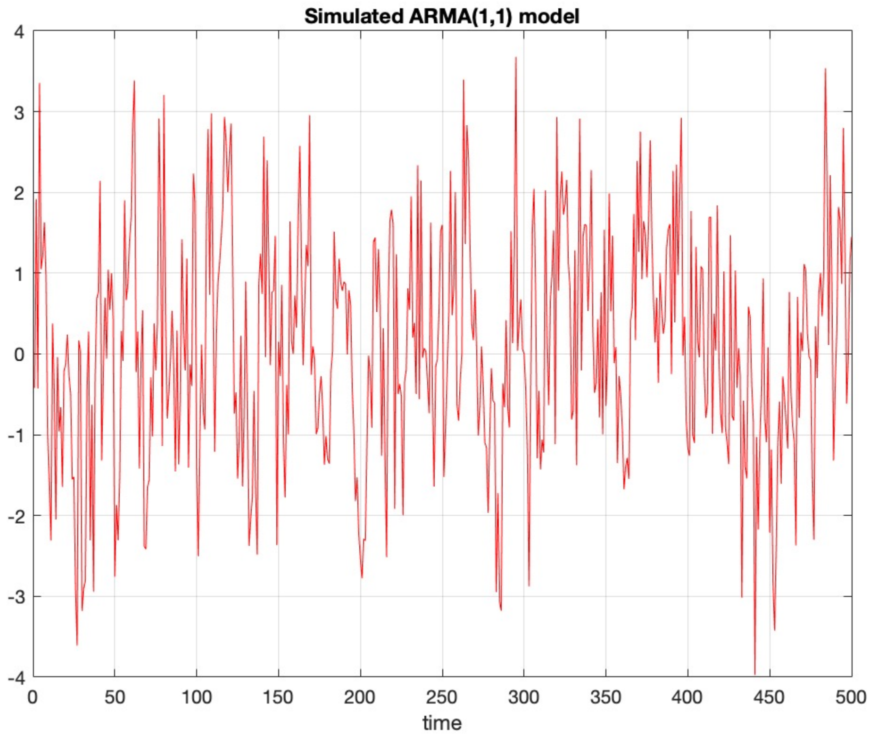
- Autocorrelation function
  - The ACF of an ARMA(p,q) process exponentially decreases to *0* when $h \to +\infty$ from order $q+1$

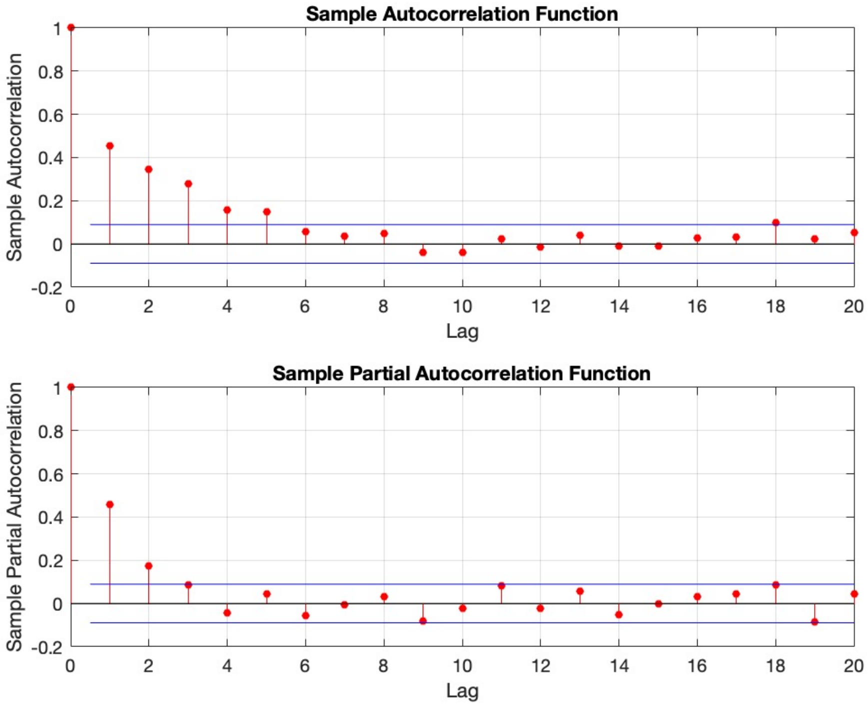
- Partial autocorrelation function
  - No special properties

- Order selection of an ARMA(p,q) process
  - *There are no such simple rules for selecting the p and q orders from the ACF and PACF plots*

ARMA(1,1) process example:

$$y_t = 0.8y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$$





No simple rules for selecting the p and q orders from the ACF and PACF plots

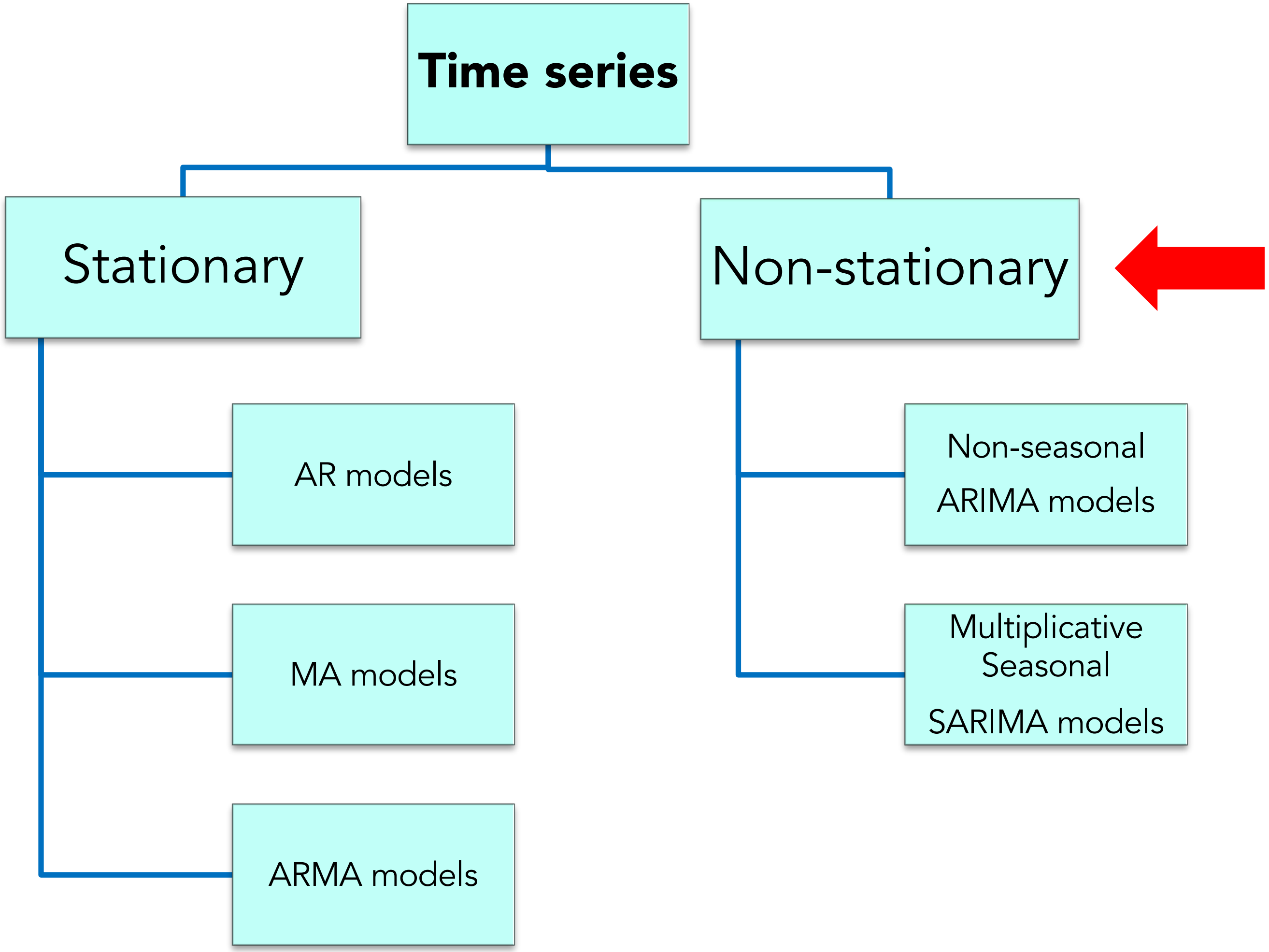
AR(p), MA(q) and ARMA(p,q) processes
Summary of ACF and PACF properties

- An AR(p) process has PACF $\alpha_y(h) = 0$ for $|h| > p$ and $\alpha_y(p) = \phi_p$

- An MA(q) process has ACF $\rho_y(h) = 0$ for $|h| > q$

- For ARMA(p, q) processes, there are no such simple rules for selecting the orders of ARMA(p, q) processes from its ACF or PACF

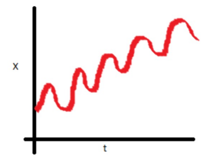
Families of ARIMA models



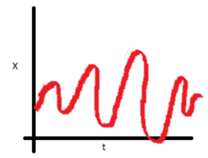
Identifying stationary/non-stationary time series

- Stationary time series:
  - is roughly horizontal
  - has constant variance
  - Has no trend nor seasonality
  - has no patterns predictable in the long-term
  - its ACF drops to zero relatively quickly

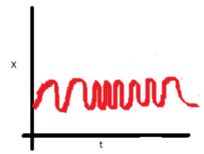
- Non-stationary time series
  - has trend and seasonality
  - its ACF decreases slowly
  - the ACF value at lag 1 is often large and positive



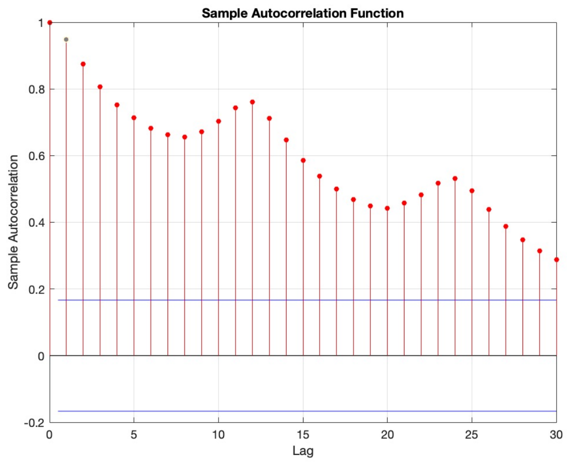
Non-Stationary series



Non-Stationary series



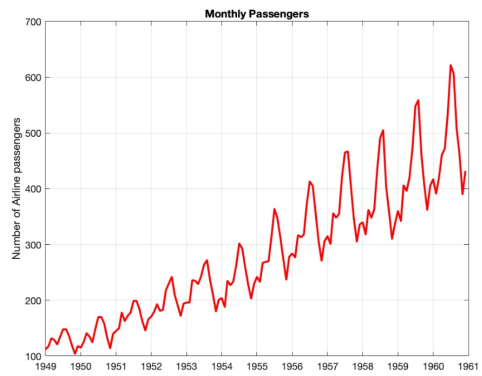
Non-Stationary series



Non-stationary time series: standard decomposition model

- Recall the standard decomposition model of a non-stationary process y_t

$$y_t = T_t + S_t + x_t$$



  - T_t is a trend-cycle component
  - S_t is a seasonality component
  - x_t is a stationary random component

- Since the Box-Jenkins methodology is for stationary models only, it is first required to detrend and deseasonalize the nonstationary series by using one of the two methods below
  - Estimate *(by linear regression)* and then remove a deterministic trend and seasonality
  - Difference the time series

Note that Box-Jenkins seemed to prefer the differencing method while several others prefer the deterministic trend removal method

Operator of lag-T

- Let Δ_T be the operator of *lag-T*

$$\Delta_T y_t = (1 - L^T)^1 y_t = y_t - y_{t-T}$$

A lag-T differencing of order 1 is applied to the time series

- Applying Δ_T *d* times in a successive way to a time series

$$(\Delta_T)^d y_t = (1 - L^T)^d y_t$$

A lag-T differencing of order d is applied to the time series

Lag-1 differencing to remove polynomial trend and achieve stationarity

- Let y_t be a time series with a polynomial trend of order *k* :

$$y_t = \sum_{i=0}^{k} \beta t^i + x_t$$

- Applying the operator of lag-1 Δ_1 to the time series

$$\Delta_1 y_t = y_t - y_{t-1}$$

  - Then, the lag-1 differenced time series will have a polynomial trend of order *k-1*
  - lag-1 difference reduces by 1 the degree of a polynomial trend

⇒ ***Applying successive Lag-1 differencing removes trend***

How to choose the order d of lag-1 differencing ?

- To remove deterministic trends
  - Apply lag-1 differencing of order d on the time series
    $$(\Delta_1)^d y_t = (1-L)^d y_t$$
  - When d=1, we have the simple first difference of the time series
    $$\Delta_1 y_t = (1-L)y_t = y_t - y_{t-1}$$
  - When d=2, we have the double difference of the time series
    $$(\Delta_1)^2 y_t = (1-L)^2 y_t = y_t - 2\, y_{t-1} + y_{t-2}$$

- How to choose the order d of lag-1 differencing ?
  - In practice $d = \{0\,;1;2\}$
    - d=0: no differencing (no trend)
    - d=1: perform differencing once (*to remove linear trend*)
    - d=2: double-differencing (*to remove quadratic trend*)

Lag-s differencing to remove seasonality trend and achieve stationarity

- Let y_t be a time series with a trend T_t and a season pattern S_t of period s ($S_{t+s} = S_t$):

$$y_t = T_t + S_t + x_t$$

- Applying the operator Δ_s to the time series

$$\Delta_s y_t = y_t - y_{t-s} = (T_t - T_{t-s}) + (S_t - S_{t-s}) + (x_t - x_{t-s})$$

$$\Delta_s y_t = (T_t - T_{t-s}) + (x_t - x_{t-s})$$

  - Then, the lag-s differentiated time series does not present any more seasonal pattern

⇒ Applying lag-s differencing removes a seasonal pattern of period s

How to choose the order D of lag-s differencing?

- To remove deterministic seasonality
  - Apply lag-s differencing of order D on the time series

$$(\Delta_s)^D y_t = (1 - L^s)^D y_t$$

- How to choose the order D of lag-s differencing?
  - In practice $\mathrm{D} = \{0\,;1\}$
    - D=0: no differencing (no seasonality)
    - D=1: perform differencing once (*to remove seasonality*)

- Example
  - if s=12, we have the lag-12 differencing of the series as *(for monthly time series data and annual seasonality for example)*

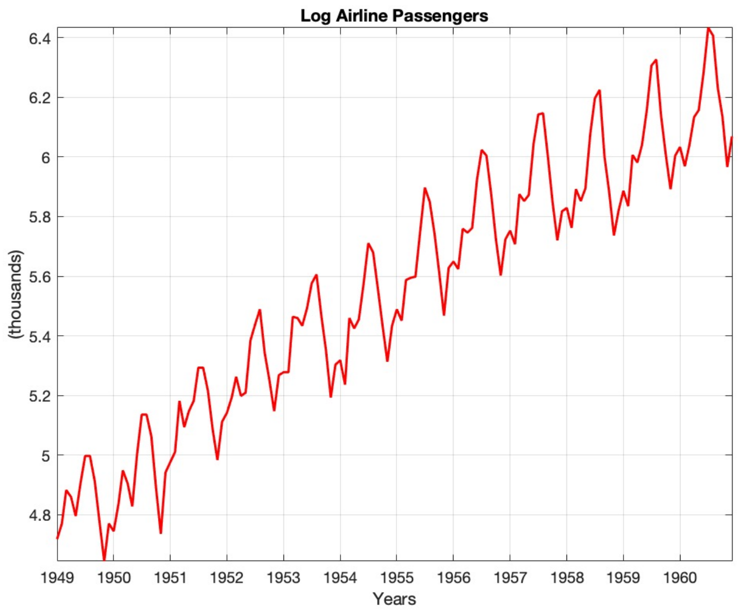
$$\Delta_{12} y_t = (1 - L^{12}) y_t = y_t - y_{t-12}$$

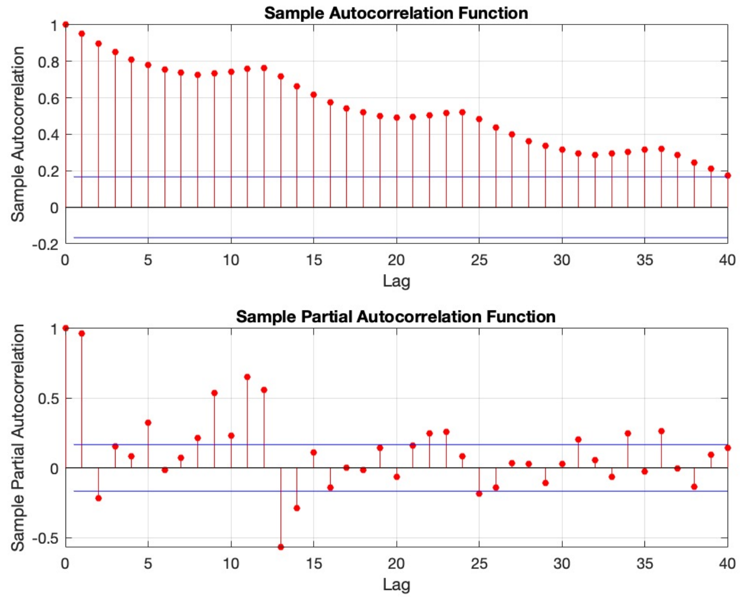
Differencing in practice

- Advantage:
  - easy to understand
  - allows forecast since we can forecast $\Delta_s y_t$ and then go back to y_t
- In practice:
  - Start by removing the seasonality trend by applying Δ_s
  - Plot the deseasonalized time series and check whether it seems stationary
  - If it does not visually seem stationary, apply then again Δ_1
  - Plot the deseasonalized and differenced time series and check whether it now seems stationary
  - If not, apply again Δ_1, but try to keep small the value for the number of differencing times

Beware of over-differencing

Converting nonstationary to stationary time series by differencing - Example





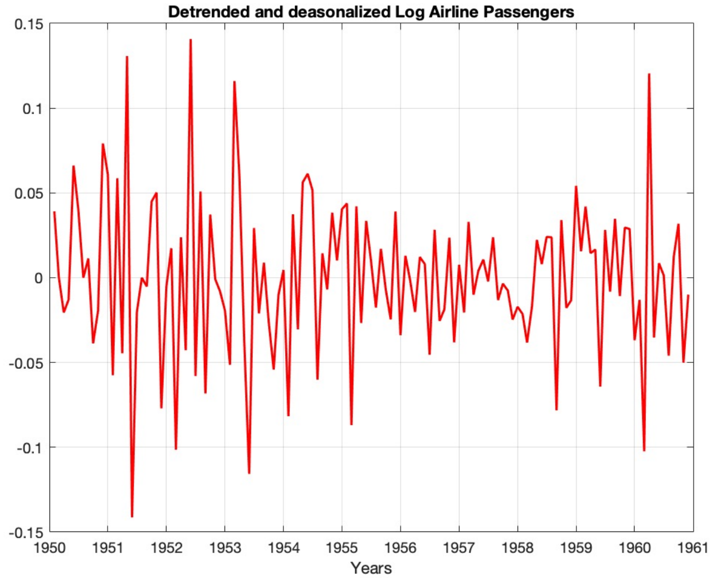
The data look nonstationary, with a linear trend and seasonal periodicity

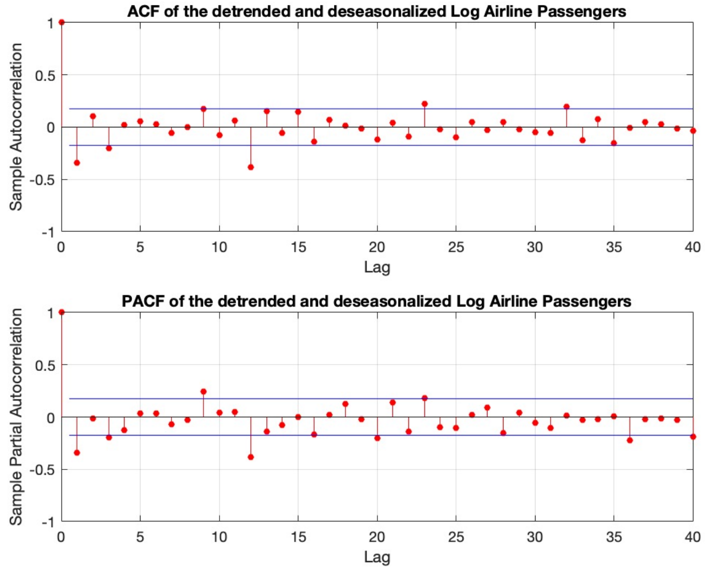
The ACF does not die out quickly and shows a cyclical pattern of period 12.

This also points to nonstationarity in the time series

Converting nonstationary to stationary time series by differencing - Example

A seasonal differences of length 12 has been applied. The linear trend has been removed by first-differencing the data.





The differenced series appears now much more stationary

Although the sample ACF and PACF of the differenced series still show significant autocorrelation at certain lags, they seem correspond to a stationary process.

The remaining autocorrelation could be captured by an ARMA model

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

ARIMA models for non seasonal time series data

- The general non-seasonal model is known as ARIMA(p,d,q):

$$\underbrace{\Phi(L)}_{\text{AR part of order p}}\ \underbrace{(1-L)^d}_{\text{Lag-1 differencing of order d}}\ (y_t - c) = \underbrace{\Theta(L)}_{\text{MA part of order q}}\ \varepsilon_t$$

AR part of order p

Lag-1 differencing of order d

MA part of order q

- y_t is an ARIMA(p,d,q) model if $(1-L)^d(y_t - c)$ is an ARMA(p,q) model

$$(1-L)^d(y_t - c) = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Understanding ARIMA(p,d,q) model orders
Example

Consider the following ARIMA(2,1,1) model:

$$\underbrace{(1-\phi_1 L^1-\phi_2 L^2)}_{\substack{\text{AR part} \\ \text{of order p=2}}}\ \underbrace{(1-L)}_{\substack{\text{Lag-1 differencing} \\ \text{of order d=1}}}\ (y_t - c) = \underbrace{(1+\theta_1 L)}_{\substack{\text{MA part} \\ \text{of order p=1}}}\ \varepsilon_t$$

The model includes all consecutive AR and MA lags from 1 through their respective orders p and q

Understanding ARIMA models

- The general non-seasonal model is known as ARIMA(p,d,q):

$$\Phi(L)(1-L)^d(y_t - c) = \Theta(L)\varepsilon_t$$

- The intercept *c* of the model and the differencing order *d* have an important **effect on the long-term forecasts**:
  - $c=0$ and $d=0 \Rightarrow$ long-term forecasts will go to 0
  - $c=0$ and $d=1 \Rightarrow$ long-term forecasts will go to constant $\neq 0$
  - $c=0$ and $d=2 \Rightarrow$ long-term forecasts will follow a straight line
  - $c\neq 0$ and $d=0 \Rightarrow$ long-term forecasts will go to the mean of the data
  - $c\neq 0$ and $d=1 \Rightarrow$ long-term forecasts will follow a straight line
  - $c\neq 0$ and $d=2 \Rightarrow$ long-term forecasts will follow a quadratic trend

Special ARIMA models

- ARIMA(0,1,0) = random walk
- ARIMA(0,1,1) without constant = simple exponential smoothing
- ARIMA(0,2,1) without constant = linear exponential smoothing
- ARIMA(1,1,2) with constant = damped-trend linear exponential smoothing

How to choose ARIMA orders (p, d, q) in practice ?

- Two situations can occur, depending on your goal:
  - obtain an understanding of the model
  - obtain a very good forecast

- General tips
  - Start by differencing the series if needed, in order to obtain something visually stationary
  - Look at the ACF and PACF plots and identify possible model orders
  - Estimate several models and select the best one by using model selection criteria such as AIC or BIC

SARIMA models for seasonal time series data

- The multiplicative seasonal model is known as SARIMA(p,d,q) $\times$(P,D,Q)$_s$:

$$(1-\phi_1 L^1-\cdots-\phi_p L^p)(1-\Phi_1 L^P-\cdots-\Phi_P L^{sP})(1-L)^d(1-L^s)^D(y_t-c)=(1+\theta_1 L+\cdots+\theta_1 L^q)(1+\Theta_1 L^Q+\cdots+\Theta_Q L^{sQ})\varepsilon_t$$

  - p is the number of non-seasonal AR terms
  - d is the order of non-seasonal first difference (lag-1)
  - q is the number of non-seasonal MA terms
  - s is the number of time periods for a season
  - P is the time lag seasonal AR (SAR) terms
  - D is the order of seasonal differences (lag-s)
  - Q is the time lag seasonal MA (SMA) terms

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Understanding SARIMA(p,d,q) × (P,D,Q)$_s$ model orders Example

Consider the following SARIMA (2,1,1) × (2,1,1)$_{12}$ model:

$$\underbrace{(1-\phi_1 L^1-\phi_2 L^2)}_{\text{AR part of order p=2}}\underbrace{(1-\Phi_{12} L^{12}-\Phi_{24} L^{24})}_{\text{SAR part of order P=2}}\underbrace{(1-L)}_{\text{Lag-1 differencing of order d=1}}\underbrace{(1-L^{12})}_{\text{Lag-12 differencing of order D=1}}(y_t-c)=c+\underbrace{(1+\theta_1 L)}_{\text{MA part of order q=1}}\underbrace{(1+\Theta_{12} L^{12})}_{\text{SMA part of order Q=1}}\varepsilon_t$$

- The period of the season is s=12
- The model includes all consecutive AR and MA lags from 1 through their respective orders p and q
- The lags of the SAR and SMA polynomials are consecutive multiples of the period (s=12) from 12 through their respective specified order P and Q, times 12

How to choose SARIMA orders $(P,D,Q)_s$ in practice

- The seasonal part of an AR or MA model can be seen in the seasonal lags of the ACF and PACF

- Examples
  - an $SARIMA(0,0,0)(1,0,0)_{12}$ will show
    - a spike at lag 12 in the PACF, and no other significant spikes
    - an exponential decay in the seasonal lags of the ACF
  - an $SARIMA(0,0,0)(0,0,1)_{12}$ will show
    - a spike at lag 12 in the ACF, and no other significant spikes
    - an exponential decay in the seasonal lags of the PACF

H. Garnier

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

How to choose all the orders (p,d,q) × (P,D,Q) of a SARIMA model in practice

- Use visual inspection to observe the trend and seasonality
  - Look at the ACF and PACF plots and identify possible orders

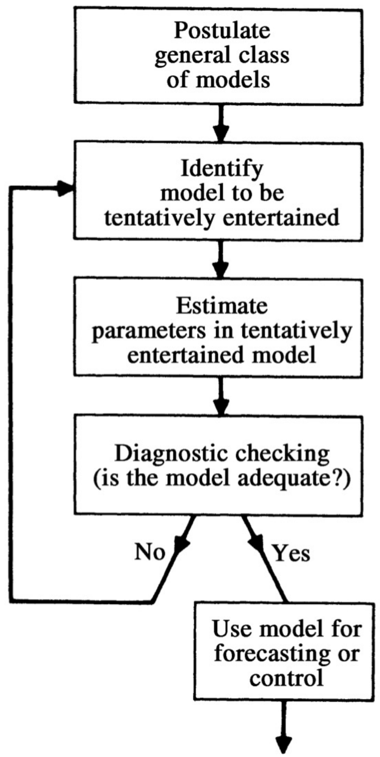
- General tips
  - Use differencing to remove the trend and seasonality
  - Keep the orders simple
    - According to Box-Jenkins
      - *"the maximum value of each SARIMA model orders p, d, q, P, D, Q is 2"*
    - According to Robert Nau, Duke University
      - *"In most cases, either p or q is zero and p+q ≤ 3"*
    - $\mathbf{d = \{0\,;1;2\},\ \ p + q \leq 3}$
    - $\mathbf{D = \{0\,;1\}}$, **P=** $\mathbf{\{0\,;1\}}$, **Q=** $\mathbf{\{0\,;1\}}$
    - Standard value for the period s =12 for an annual seasonality if monthly time series data

The Box-Jenkins methodology

- The Box-Jenkins methodology refers to a set of stages for identifying, fitting, and checking ARIMA models for time series data

- The basis of Box-Jenkins approach to modeling time series consists of three main stages:
  1. Identification
  2. Estimation
  3. Diagnostics

- Forecasts follow directly from the form of fitted model



Stages in the iterative approach to model building.

UNIVERSITÉ DE LORRAINE POLYTECH NANCY

The Box-Jenkins methodology:
Estimation and model selection

- Once the orders (p, d, q) are selected, Maximum Likelihood Estimation (MLE) through optimization algorithms can be used to estimate the model parameters

- MLE cannot be used to choose orders(p,d,q)
  - the larger (p, d, q) ⇒ the larger the number of parameters ⇒ the more flexible the model ⇒ the larger the likelihood
  - MLE should be penalized by the complexity of the model (≃ number of parameters to be estimated)

- *Some model selection criteria can be used. The idea is to test a range of possible model candidates and to compute the criteria for each model structure tested*

The Box-Jenkins methodology:
Model selection

- Let $L(\boldsymbol{\theta})$ denote the value of the maximized likelihood objective function for a model with a total of n_p parameters fitted to N data points

- Information criteria are likelihood-based measures of model fit that include a penalty for complexity, specifically, the number of parameters n_p

- Different information criteria are distinguished by the form of the penalty, and can favor different models
  - Akaike information criterion (AIC) : $-2\log L(\boldsymbol{\theta}) + 2\, n_p$
  - Bayesian information criterion (BIC) : $-2\log L(\boldsymbol{\theta}) + n_p \log(N)$

- When you compare values for a set of model candidates, smaller values of the criterion (AIC or BIC) indicate a better, more parsimonious model

Occam's Razor Rule

designates by metaphor the opportunity to "cut off", as with a razor, the superfluous assumptions of a theory

"Among models with similar performance,

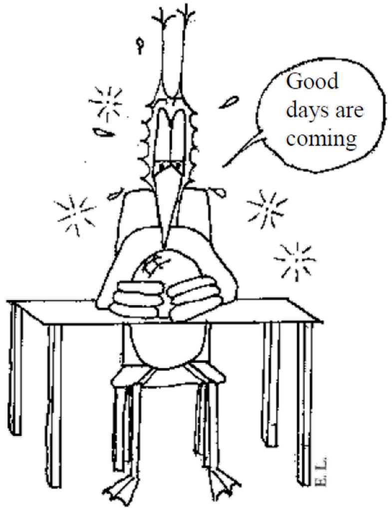
select the model that describes data with fewest parameters"

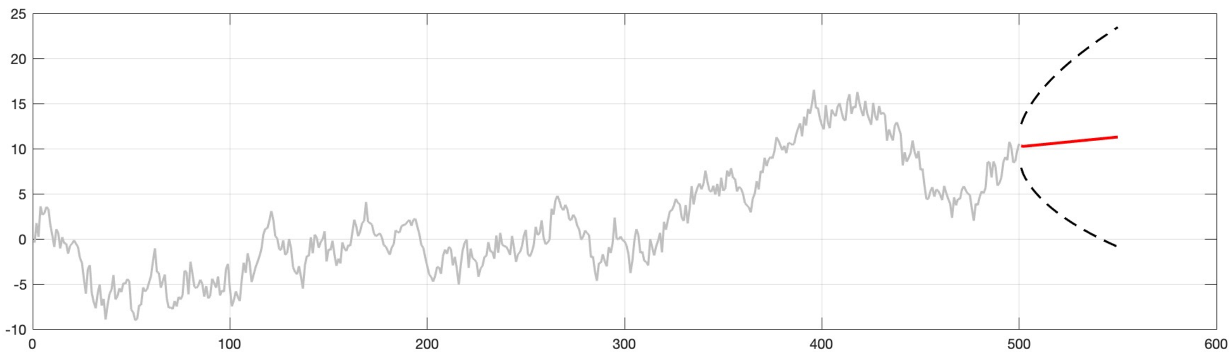
The Box-Jenkins methodology:
Forecasting

- It is impossible to forecast without error

- The good engineer should
  - forecast *what can be forecast*

  AND
  - provide *uncertainty* intervals

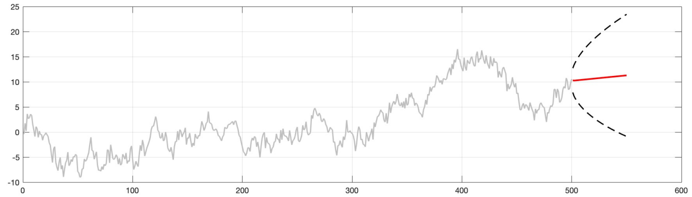




Confidence intervals

- Assuming that the *residuals are normally distributed*, we can usefully assess the accuracy of a forecast by using $\sqrt{\text{MSE}}$ as an estimate of the error

  where $\text{MSE} = \frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2$



- An approximate prediction interval for the next observations is

$$\hat{y}_{t+1} \pm z\sqrt{\text{MSE}}$$

  where z is a quantile of the normal distribution. Typical values used are given in the table

z	Probability
1.282	0.80
1.645	0.90
1.960	0.95
2.576	0.99

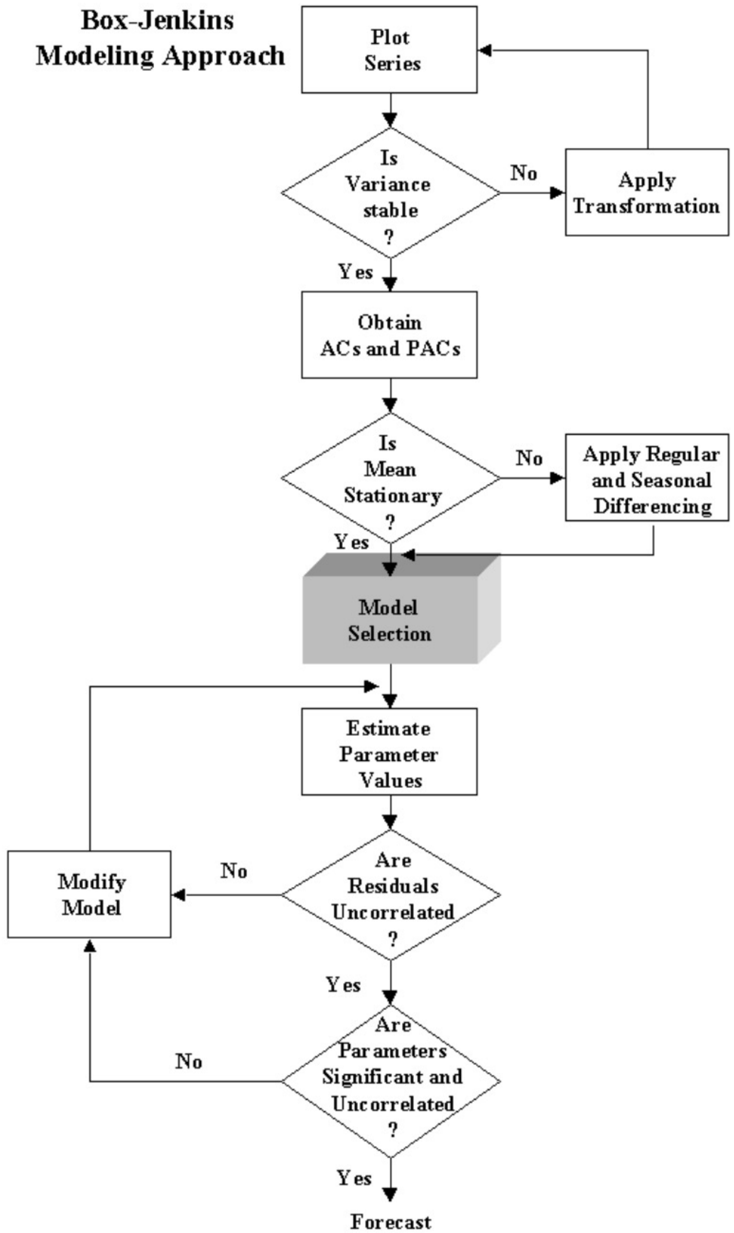
- This enables, for example, 95% or 99% confidence intervals to be set up for any forecast

The Box-Jenkins methodology in details

I. Identification
   1. Data preparation
      a. Transform data to stabilize variance (apply logarithm, etc)
      b. Differencing data to obtain stationary series
   2. Model selection
      a. Examine data to identify potential models
      b. Examine ACF, PACF
      c. Use automatic search methods

II. Estimation
   1. Estimate parameters in potential models
   2. Select best model using suitable information criteria (AIC, BIC,…)

III. Diagnostics
   1. Check ACF/PACF of residuals
   2. Are the residuals white noise?
   3. Do more statistical tests of residuals



home.ubalt.edu/ntsbarsh/stat-data/BJApproach.gif

Two case studies

Let us apply the Box-Jenkins methodology to

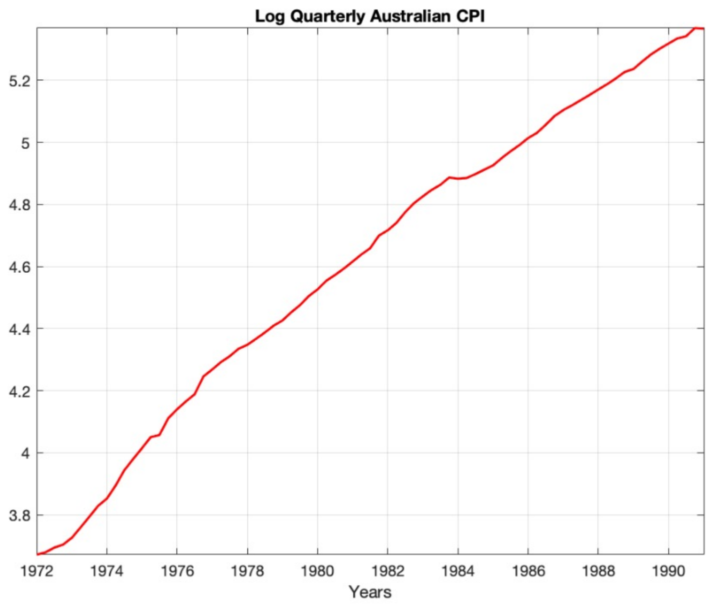
- ARIMA model estimation and forecast - Australian Consumer Price Index
- SARIMA model estimation and forecast - International airline passenger data

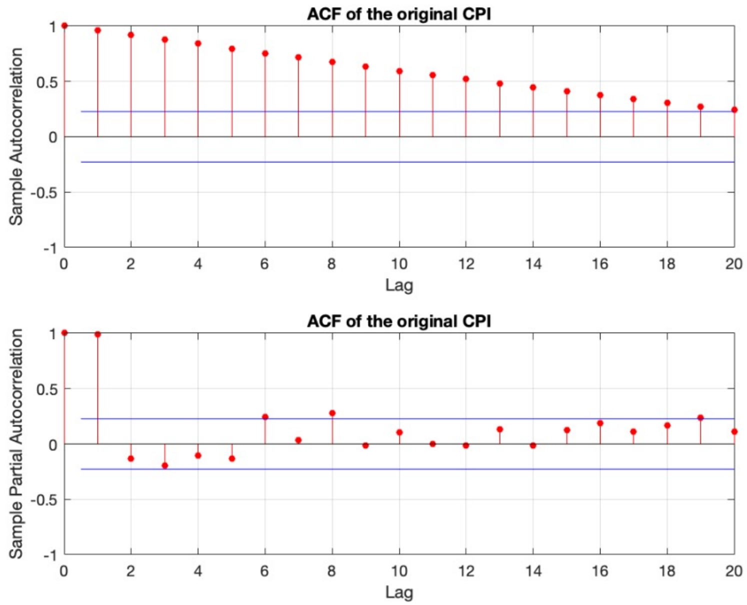
For each case study

1. Plot the time series and check whether the variance needs to be stabilized
2. Check whether it is stationary. Does it show trends and seasonality?
3. Apply the differencing method to remove possible trend and seasonal pattern
4. Specify the period of the seasonal pattern (if any), the degree of the polynomial trend.
5. Check whether the differenced series seems stationary? Does it look like a white noise?
6. If not, determine the best ARMA model structure for the time series and estimate the full model form
7. Use your best model to forecast the time series over the next 5 years

First case study – Australian Consumer Price Index (CPI)
Step 1 – Identification

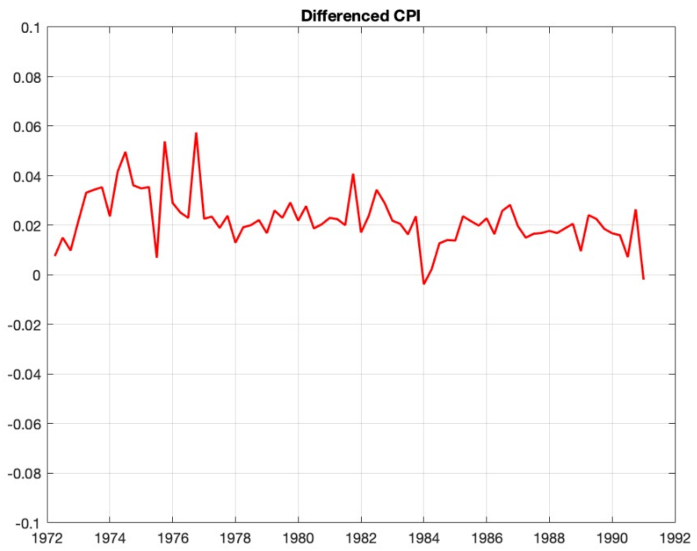
- Data available are the logarithm Australian CPI. The variance of the log CPI remains constant over time. There is no need for further transformation

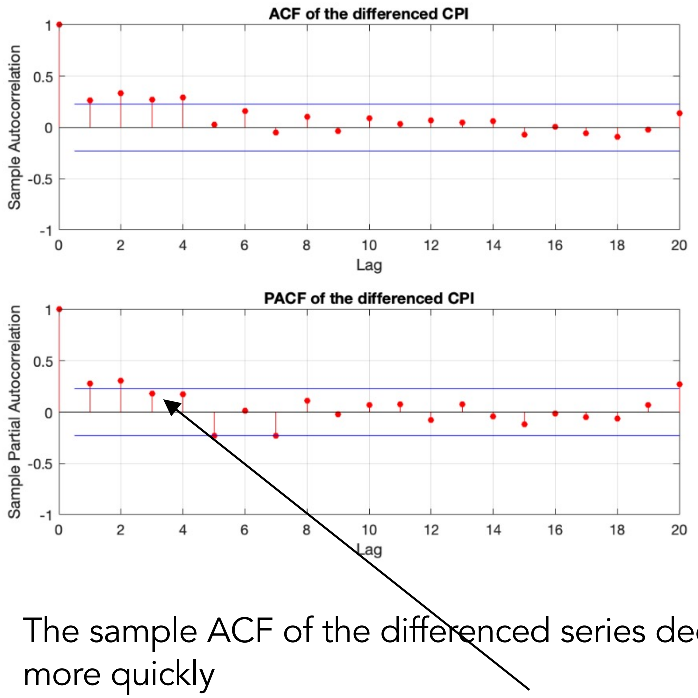




- From the time plot, the time series is nonstationary, with a clear upward trend, also noticeable from the slow decrease of the ACF
- We need to remove the linear trend by first differencing the data

Step I.2 – Model selection
Differencing data to obtain stationary data. Observe its ACF and PACF





The differenced series appears now stationary although not zero mean

The sample ACF of the differenced series decays more quickly

The sample PACF cuts off after lag 2. This behavior is consistent with a second-order autoregressive AR(2) model

Step 2 – Estimation
Estimate parameters of the chosen model structure

- The following ARIMA(2,1,0) model has been selected as a potential model:

Non-seasonal AR(2) First difference No non-seasonal MA part in the model

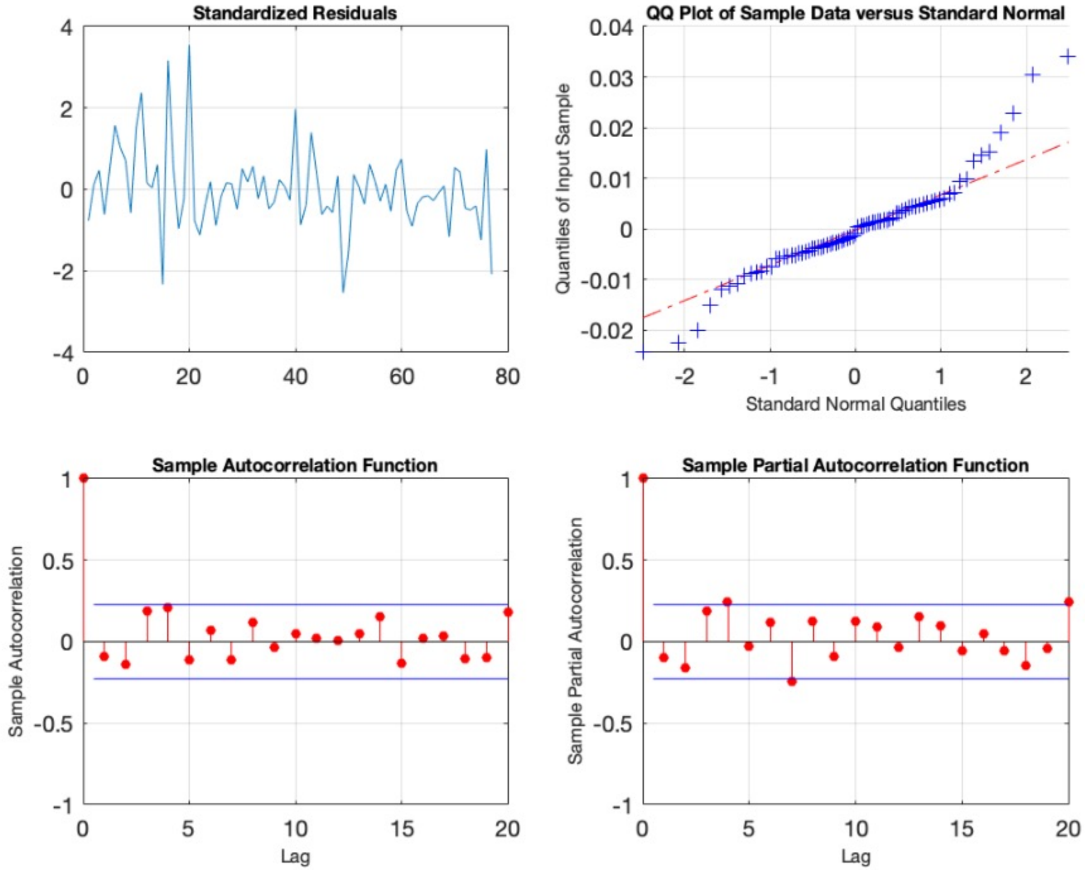
$$\left(1-\phi_1 L^1-\phi_2 L^2\right)(1-L)(y_t-c)=\varepsilon_t$$

- The constant c and two AR (ϕ_1 and ϕ_2) model parameters have been estimated by Maximum Likelihood through optimization algorithms (see **estimate** in Matlab)

ARIMA(2,1,0) Model (Gaussian Distribution):

	Value	StandardError
Constant	0.010072	0.0032802
AR{1}	0.21206	0.095428
AR{2}	0.33728	0.10378
Variance	9.2302e-05	1.1112e-05

Step 3 – Diagnostics
Analysis of the residuals

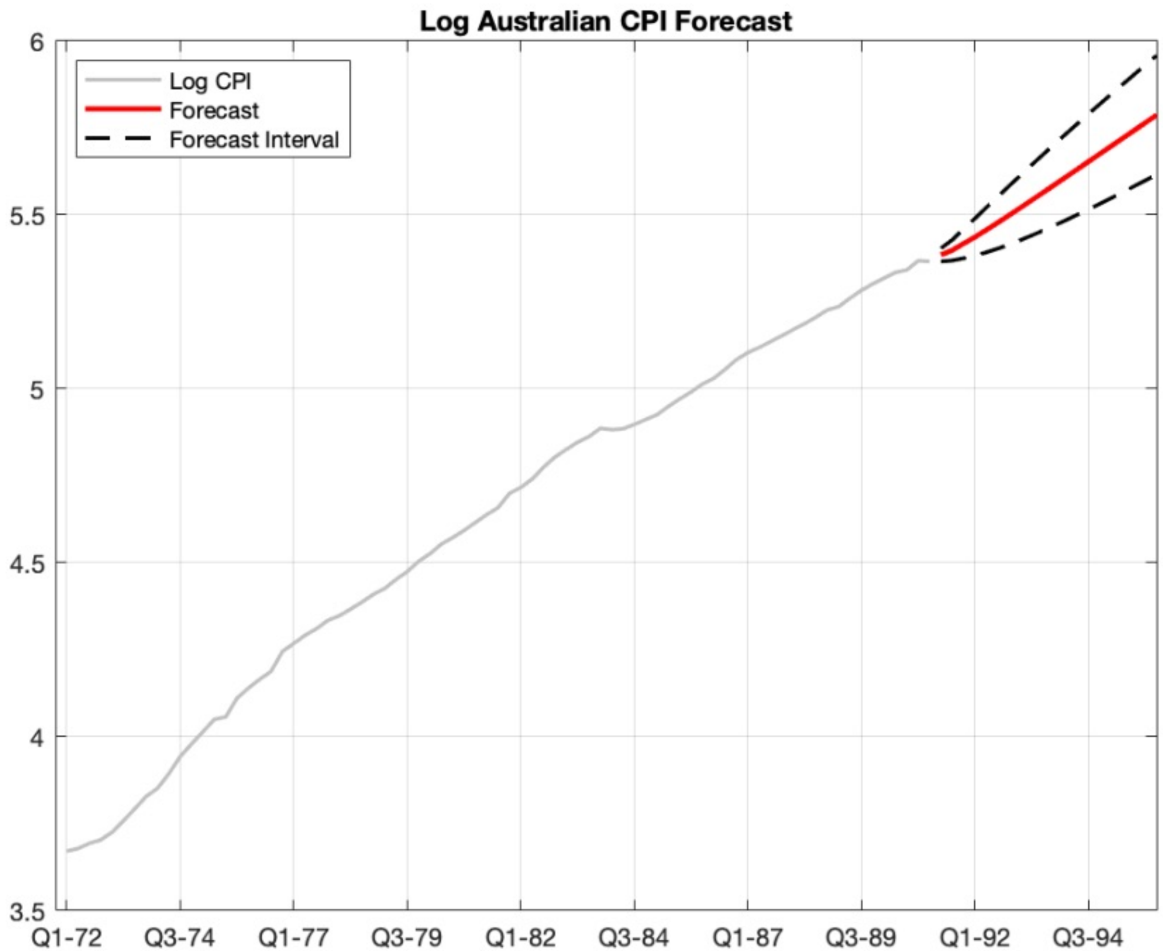


All ACF and PACF coefficients lie within the limits, indicating that the residuals are white

(more precisely, the residuals cannot be distinguished from white noise)

Step 4 – Forecasting

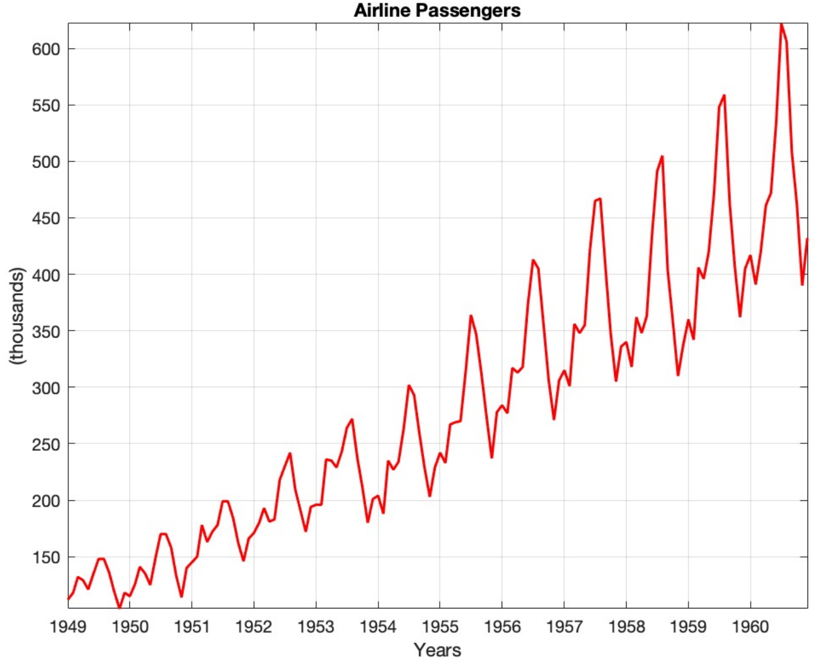
Use of the estimated model to forecast the next 5 years

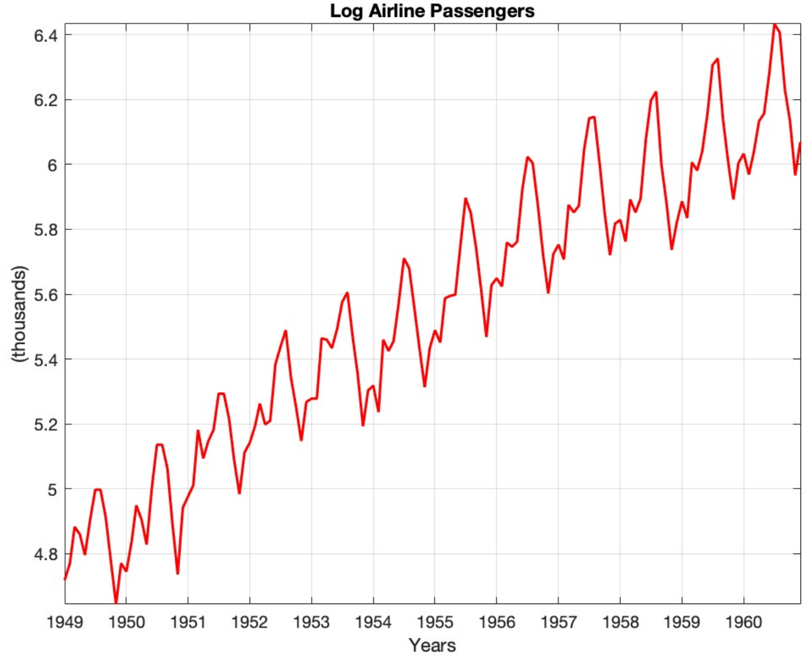


Second case study – Airline passengers

Step 1 – Identification

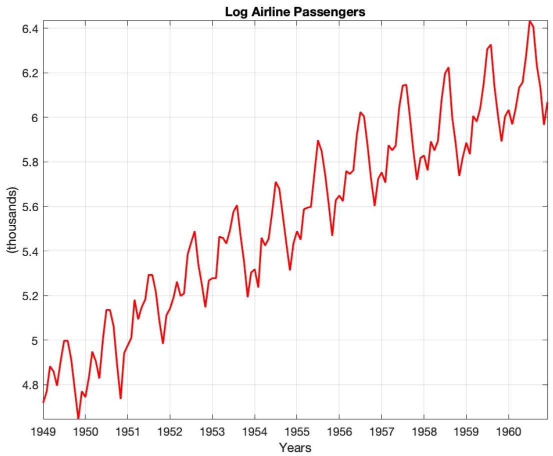
Transform data to stabilize variance by applying logarithm

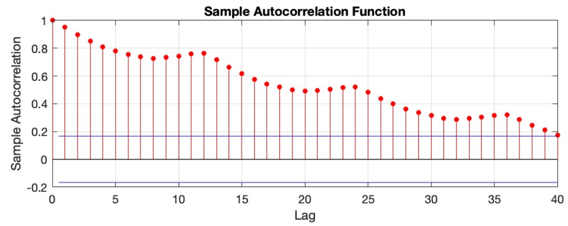


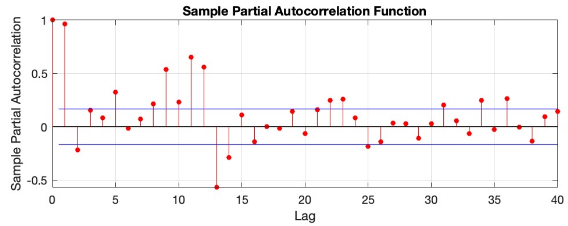


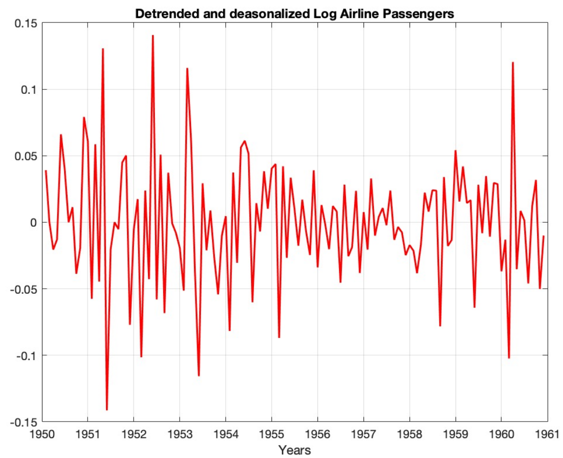
Step 1 – Identification
Differencing data to obtain stationary series

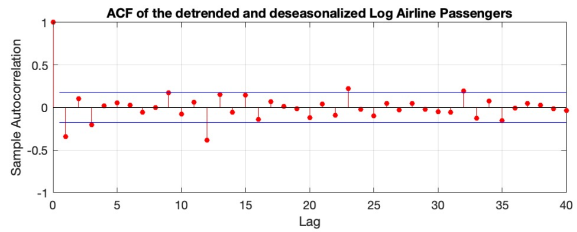
- Because these is monthly data, we use seasonal differences of length 12
- We also remove a linear trend by first differencing the data

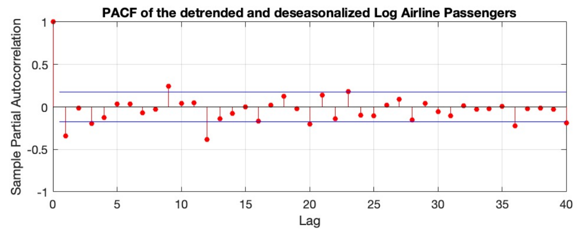




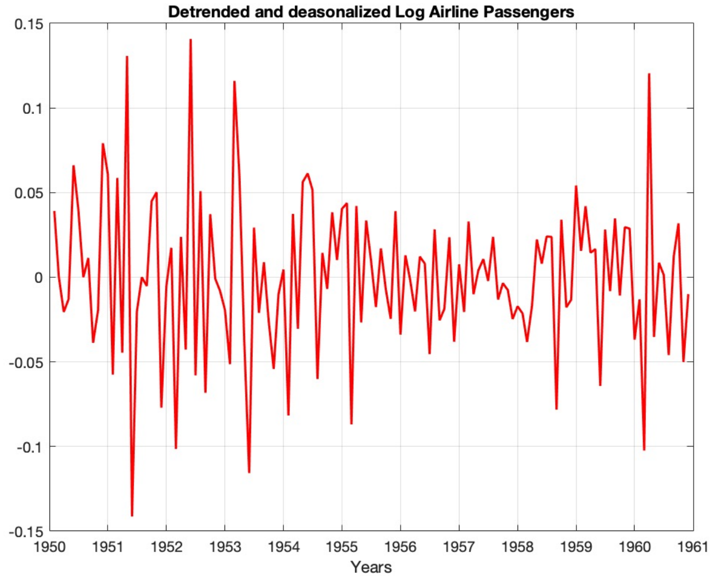


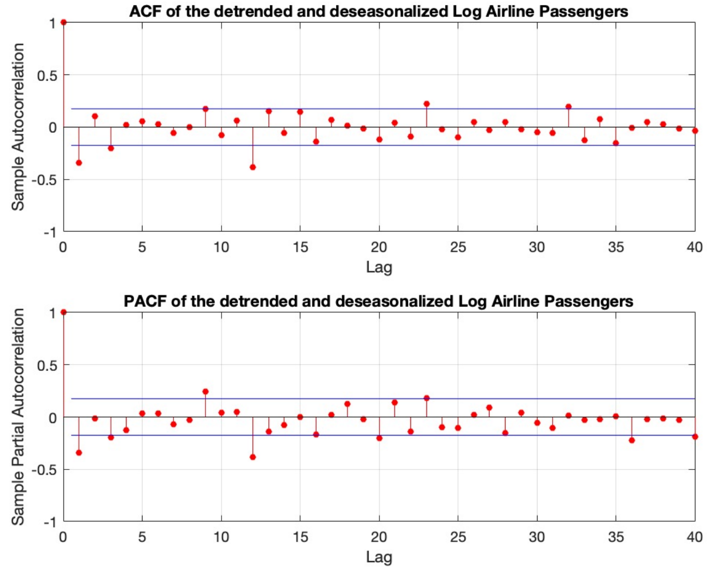






Step I.2 – Model selection
Examine data, ACF and PACF



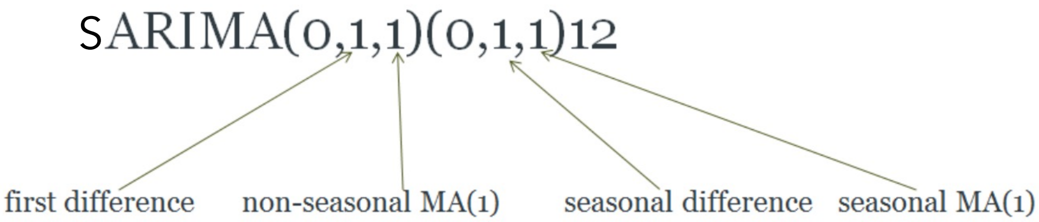


The differenced series appears now stationary

The sample ACF and PACF of the differenced series still show significant autocorrelation at lags that are multiples of 12. There is also potentially significant autocorrelation at smaller lags. It has been shown that this autocorrelation can be best captured by an SARIMA$(0,1,1) \times (0,1,1)_{12}$ model (from AIC and BIC model selection tests, not shown here)

Step 2 – Estimation
Estimate parameters of the chosen model structure

- The following SARIMA(0,1,1) × (0,1,1)$_{12}$ model has been selected as a potential model:



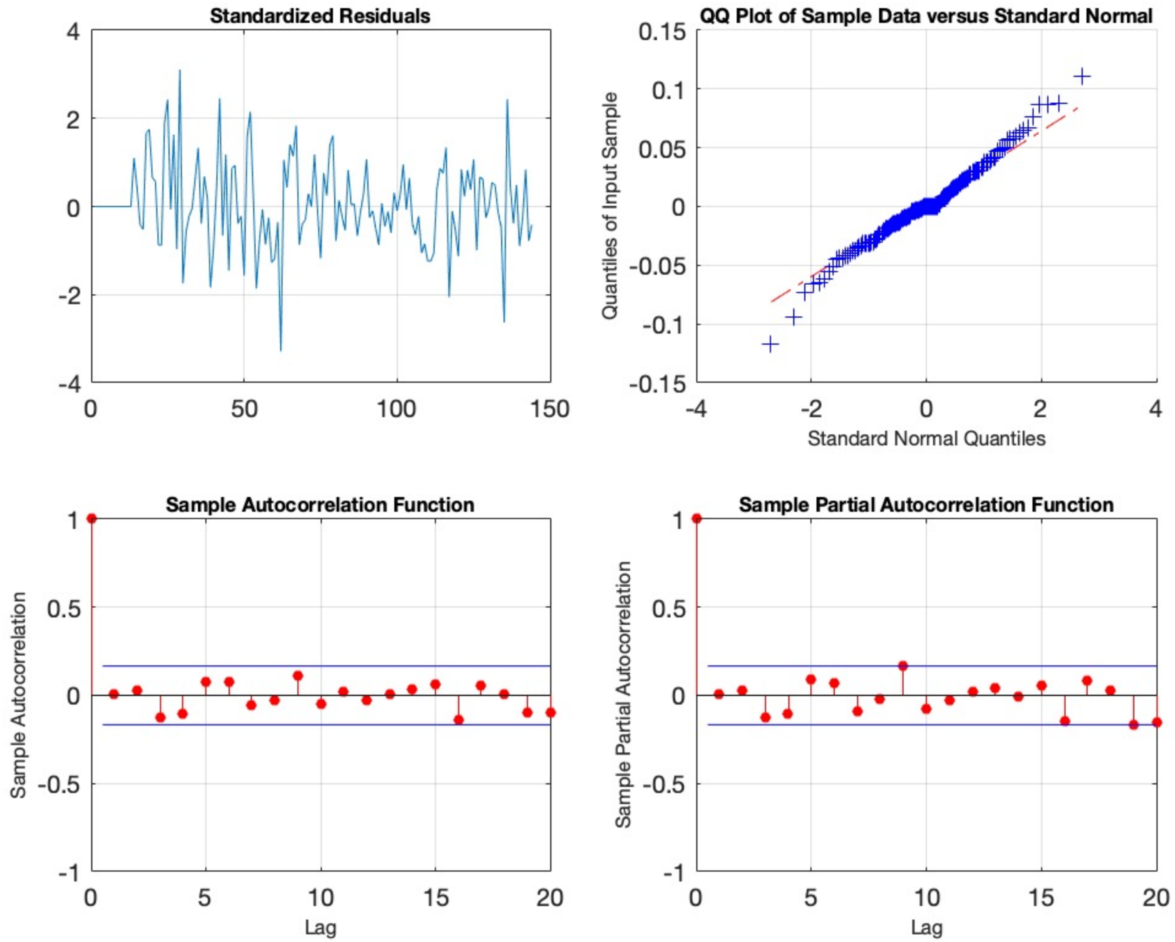
$$(1-L)(1-L^{12})y_t = (1+\theta_1 L)(1+\Theta_{12}L^{12})\varepsilon_t$$

- The MA (θ_1) and SMA (Θ_1) model parameters have been estimated by Maximum Likelihood through optimization algorithms (see **estimate** in Matlab)

ARIMA(0,1,1) Model Seasonally Integrated with Seasonal MA(12) (Gaussian Distribution):

	Value	StandardError	TStatistic	PValue
Constant	0	0	NaN	NaN
MA{1}	-0.37716	0.066794	-5.6466	1.6364e-08
SMA{12}	-0.57238	0.085439	-6.6992	2.0952e-11
Variance	0.0012634	0.00012395	10.193	2.1406e-24

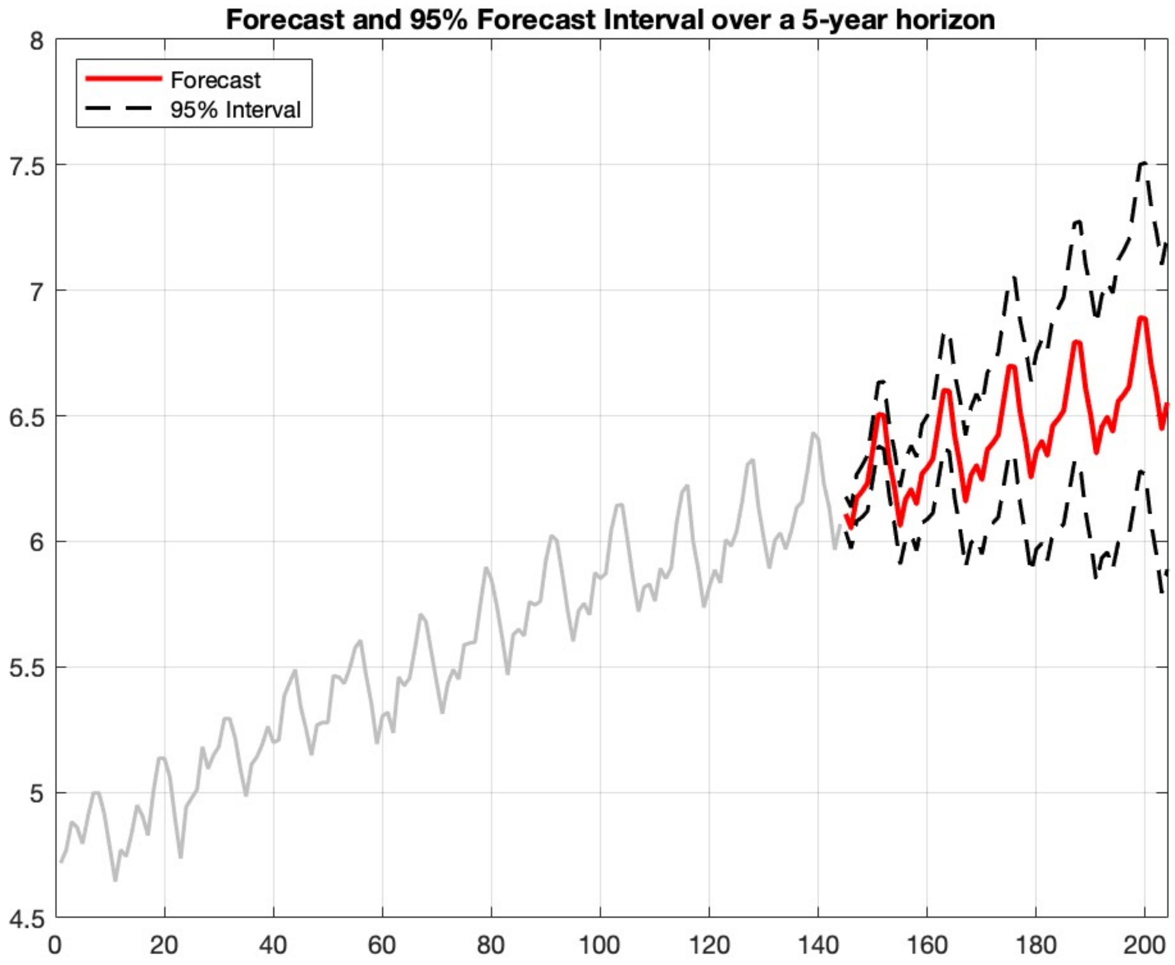
Step 3 – Diagnostics
Check residuals



All ACF and PACF coefficients lie within the limits, indicating that the residuals are white

(more precisely, the residuals cannot be distinguished from white noise)

Step 4 – Forecasting
Use of the estimated model to forecast the next 5 years



Takeaway messages

- We have just introduced basics and the core ideas behind time series analysis and forecasting
- Obviously, each problem has its own subtleties and demands special steps: proper data preparation, way of handling missing values, or defining evaluation metric satisfying some domain conditions
- It is impossible to come up with a general approach that can handle all situations
  - The Box-Jenkins method has been remarkably successful
  - More complex models and methods exist as
    - Interrupted models to include the influence of critical effects
    - GARCH models for generalized autoregressive conditional heteroscedasticity models
    - State-space models and methods
    - Recursive methods
    - Deep learning-based methods…