QBUS6840 Lecture 7

ARIMA Models (I)

Discipline of Business Analytics

The University of Sydney Business School

Recap and an example

Last half, we discussed

- Basic concepts: Forecasting problems, process of forecasting, time series components, etc.
- Time series decomposition: mainly for interpretation, can be useful for forecasting
- Time series regression: forecast based on predictors
- Exponential smoothing: can be used for both interpretation and forecasting

Example: Data Analytics in Logistic, IBM and DHL, https://www.dhl.com/content/dam/dhl/global/core/documents/pdf/glo-core-trend-report-artificial-intelligence.pdf

...By analyzing 58 different parameters of internal data, the machine learning model is able to predict if the average daily transit time for a given lane is expected to rise or fall up to a week in advance. Furthermore, this solution is able to identify the top factors influencing shipment delays, including temporal factors like departure day or operational factors such as airline on-time performance...

Table of contents

Readings

Online Textbook Sections 8.1-8.4 (otexts.org/fpp/8/); and/or BOK Ch 9 and Ch 10

Box-Jenkins Method

- A class of formal statistical time series models, often called ARIMA models
- Can capture complicated underlying patterns in the time series, rather than trend and seasonality
- Can be used as an alternative to, or in conjunction with, other forecasting techniques such as Exponential Smoothing
- Best textbook (in terms of theoretical foundation): Time Series Analysis: forecasting and control. 1st ed. 1976 (Box and Jenkins), 5th ed. 2015 (Box, Jenkins, Reinsel, Ljung).

Stationarity

The Box-Jenkins method relies heavily on the concept of stationarity

Definition

A time series process $\{Y_t\}$ is stationary if its mean, variance and covariance functions do not change over time. That is,

$$\mathbb{E}(Y_t) = \mu, \quad \mathbb{V}(Y_t) = \sigma^2,$$

and for each integer k,

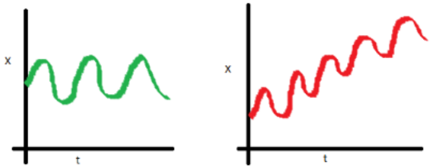
$$\mathrm{Cov}(Y_t, Y_{t-k}) = \mathrm{Cov}(Y_t, Y_{t+k}) = \gamma_k,$$

for all t.

Note: In some textbooks, this kind of stationarity is often referred to as weak stationarity.

Visually Checking Stationarity

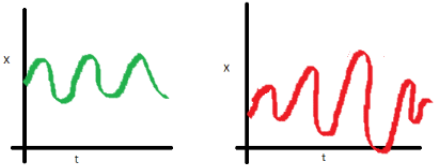
The mean of series should not be a function of time.



Credit: http://www.blackarbs.com/blog/time-series-analysis-in-python-linear-models-to-garch/11/1/2016

Visually Checking Stationarity

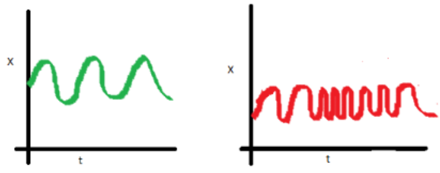
The variance of the series should not be a function of time.



Credit: http://www.blackarbs.com/blog/time-series-analysis-in-python-linear-models-to-garch/11/1/2016

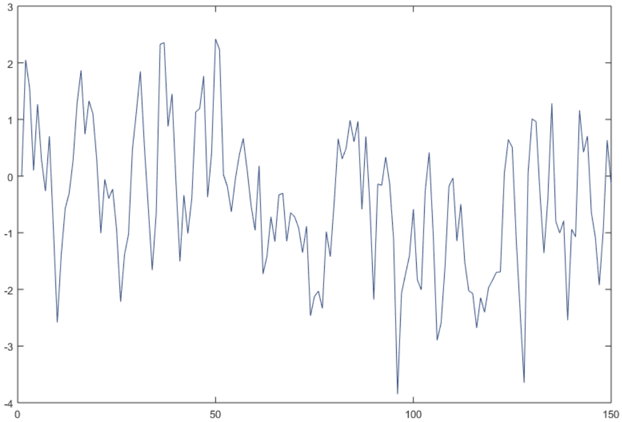
Visually Checking Stationarity

The covariance of the i-th term and the $(i + k)$-th term should not be a function of time.

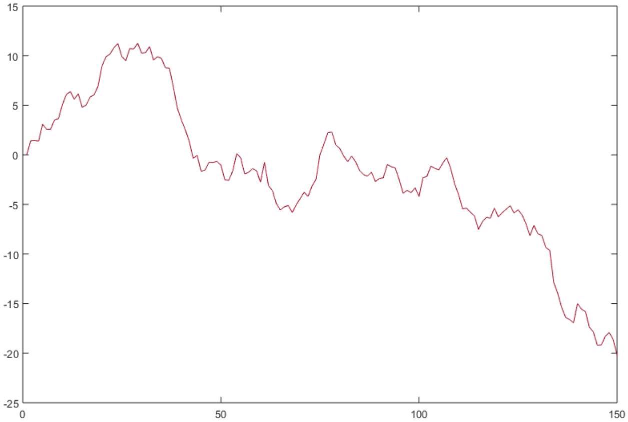


Credit: http://www.blackarbs.com/blog/time-series-analysis-in-python-linear-models-to-garch/11/1/2016

Stationarity: Illustration

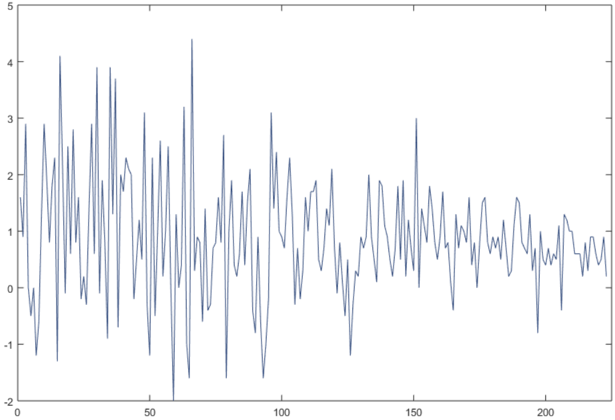


Non-stationarity: Illustration



Australian seasonally adjusted quarterly GDP growth (1959-2015)

Stationary or non-stationary?



Google stock price: Stationary or non-stationary?

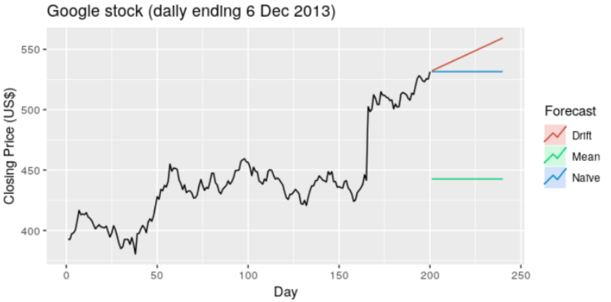


Figure 3.2: Forecasts based on 200 days of the Google daily closing stock price.

Autocorrelation function (ACF)

- Measure the correlation between observations Y_t and its lagged values Y_{t-k}, hence the name autocorrelation
- Give insights into statistical models that best describe the time series data
- Box and Jenkins advocate using the ACF plots to assess stationarity and identify a suitable model.

Autocorrelation function (ACF)

Definitions

ACF:

$$\rho_k = \frac{\mathbb{E}\left[(Y_t - \mu)(Y_{t+(-)k} - \mu)\right]}{\sqrt{\mathbb{V}(Y_t)\mathbb{V}(Y_{t+(-)k})}} = \mathrm{Corr}(Y_t, Y_{t+(-)k}).$$

Sample ACF:

$$r_k = \frac{\sum_{t=1}^{N-k}(y_{t+k} - \overline{y})(y_t - \overline{y})}{\sum_{t=1}^{N}(y_t - \overline{y})^2}.$$

What is the value of ρ_0 and r_0?

What we have done is to measure the correlation of Y_1 and Y_{1+k}, Y_2 and Y_{2+k}, etc.
k is called the lag value.

Sample ACF: Regression Explanation

- Given a time series $\{y_1, y_2, ..., y_N\}$ and a lag k, consider the following linear regression

$$y_{t+k} - \overline{y} = \gamma(y_t - \overline{y}) \quad \text{think of it as } Y = \gamma X$$

- Consider data set

X	$y_1 - \overline{y}$	$y_2 - \overline{y}$	$y_3 - \overline{y}$	$\cdots$	$y_{N-k} - \overline{y}$
Y	$y_{1+k} - \overline{y}$	$y_{2+k} - \overline{y}$	$y_{3+k} - \overline{y}$	$\cdots$	$y_N - \overline{y}$

- Then according to the least square regression solution

$$\gamma = \frac{\sum_{t=1}^{N-k}(y_t - \overline{y})(y_{t+k} - \overline{y})}{\sum_{t=1}^{N-k}(y_t - \overline{y})^2}$$

which is close to r_k.

Sample ACF: Standard errors

- Often, we want to test whether or not $H_0 : \rho_k = 0$, based on the sample ACF r_k. This is done using a t-test
- Standard error of r_k:

$$s_{r_k} = \begin{cases} \frac{1}{\sqrt{N}}, & \text{if } k = 1, \\ \frac{\sqrt{1+2\sum_{j=1}^{k-1} r_j^2}}{\sqrt{N}}, & \text{if } k > 1 \end{cases}$$

- The t-statistic is defined as

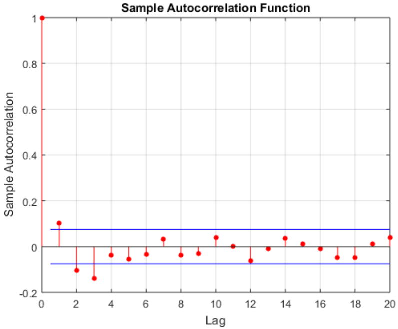
$$t_{r_k} = \frac{r_k}{s_{r_k}}$$

- Often, we reject the hypothesis $H_0 : \rho_k = 0$ if $t_{r_k} > 2$.

(Sample) ACF Plots

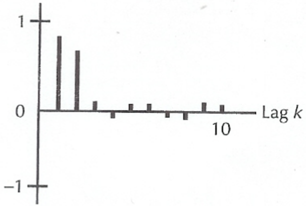
- An ACF Plot is a bar plot; the height of bar at lag k is r_k.
- We say that the plot has a spike at lag k if r_k is significantly large, i.e. its t-statistics $t_{r_k} > 2$
- The plot cuts off after lag k if there are no spikes at lags greater than k.
- We say the ACF plot dies down if the plot doesn't cut off, but decreases in a steady fashion.

(Sample) ACF Plots: Behaviour of ACFs

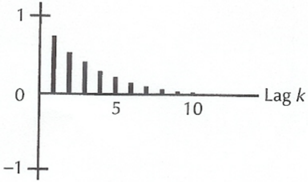


This sample ACF plot has spikes at lags 1, 2 and 3.

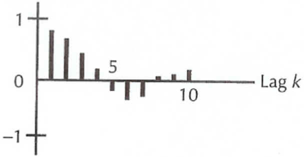
(Sample) ACF Plots: Behaviour of ACFs



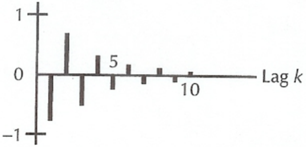
(a) Cuts off after lag 2



(b) Damped exponential dying down



(c) Damped sine-wave dying down



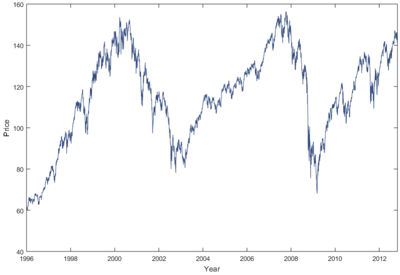
(d) Damped exponential dying down with oscillation

Assessing stationarity

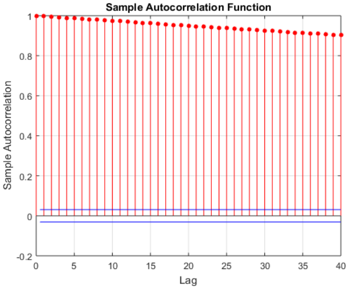
We can assess the stationarity of $\{y_t\}$ by assessing its (sample) ACF plot. In general, it can be shown that for nonseasonal time series

- If the Sample ACF of a nonseasonal time series cuts off or dies down reasonably quickly, then the time series should be considered stationary.
- If the Sample ACF of a nonseasonal time series dies down extremely slowly or not at all, then the time series should be considered nonstationary.

S&P 500 index

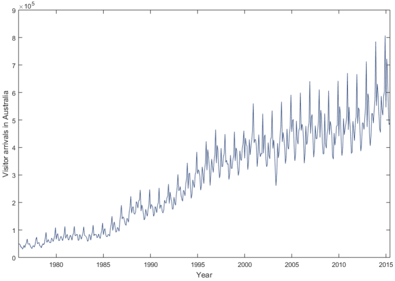


(a) Series

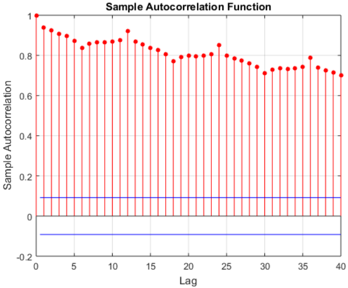


(b) ACF

Visitor arrivals in Australia

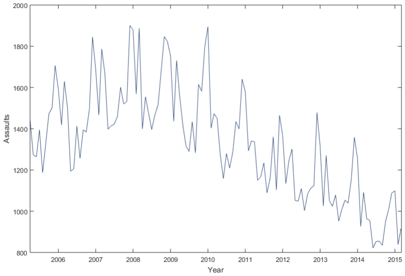


(c) Series

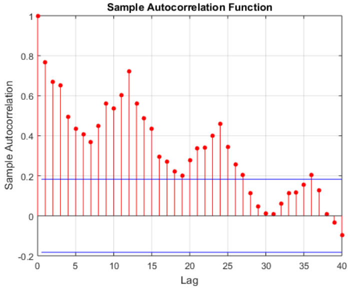


(d) ACF

Alcohol related assaults in NSW

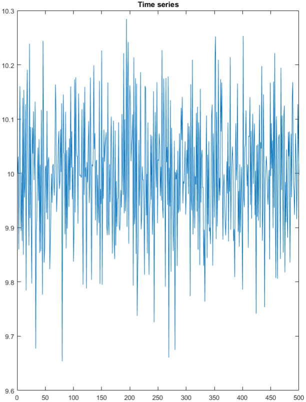


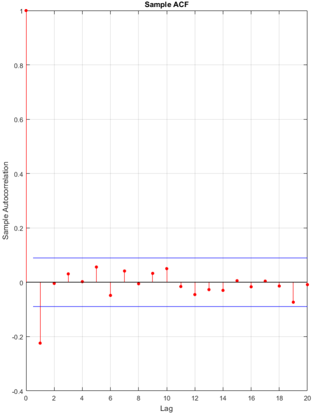
(e) Series



(f) ACF

Stationary?


Time series
10.3
10.2
10.1
10
9.9
9.8
9.7
9.6
0
50
100
150
200
250
300
350
400
450
500


Sample ACF
Sample Autocorrelation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
2
4
6
8
10
12
14
16
18
20
Lag

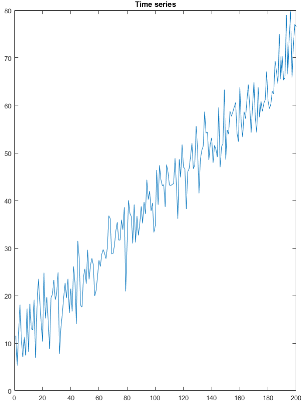
Tranforming

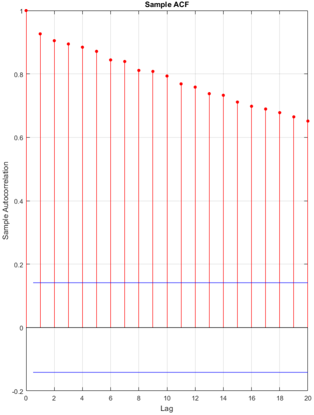
- If the ACF of a time series $\{y_1, ..., y_N\}$ dies down extremely slowly, data transformation is necessary
- Trying first order differencing is always a good way. See example `Lecture07_Example01.py`

$$z_t = y_{t+1} - y_t, \quad t = 1, ..., N-1$$

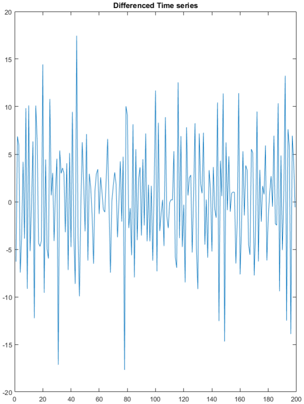
- If the ACF for the transformed data $\{z_t\}$ still dies down extremely slowly, the transformed time series should be considered nonstationary. More transformations needed
- For nonseasonal data, first or second differencing will generally produce stationary time series values.

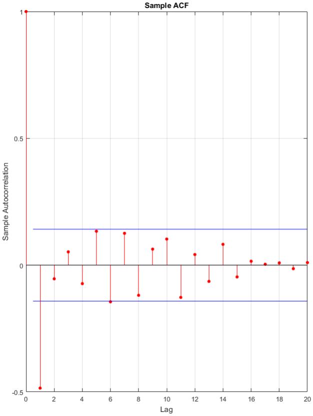
Tranforming: original time series



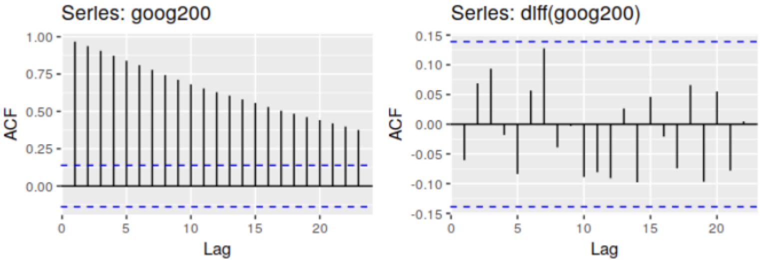


Tranforming: differenced time series





Tranforming



Partial ACF

- Partial autocorrelations measure the linear dependence of two variables after removing the effect of other variable(s) that affect both variables.
- Apart from ACF, we need Partial ACF for identifying appropriate ARIMA models.
- For example, the partial autocorrelation of 2nd order measures the effect (linear dependence) of Y_{t-2} on Y_t after removing the effect of Y_{t-1} on both Y_t and Y_{t-2}

Partial ACF

- Each partial autocorrelation could be obtained as a series of regressions of the form:

$$
\begin{aligned}
Y_t &= \rho_{10} + \rho_{11} Y_{t-1} + \varepsilon_t \\
Y_t &= \rho_{20} + \rho_{21} Y_{t-1} + \rho_{22} Y_{t-2} + \varepsilon_t \\
Y_t &= \rho_{k0} + \rho_{k1} Y_{t-1} + \rho_{k2} Y_{t-2} + \ldots + \rho_{kk} Y_{t-k} + \varepsilon_t
\end{aligned}
$$

- Note: $\rho_{11} = \rho_1$.
- The meaning of ACF coefficient ρ_k is

$$Y_t = \rho_0 + \rho_k Y_{t-k} + \varepsilon_t$$

without considering other Y_{t-k+1}, ..., Y_{t-1}.

Sample Partial ACF*

- The Sample Partial ACF at lag k is

$$r_{kk} = \begin{cases} r_1 & \text{if } k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} & \text{if } k = 2, 3, \ldots \end{cases}$$

where

$$r_{k,j} = r_{k-1,j} - r_{kk} r_{k-1,k-j} \;\; \text{for } j = 1, 2, \ldots, k-1$$

- The standard error of r_{kk} is

$$s_{r_{kk}} = \frac{1}{\sqrt{N}}$$

First Simple Process: White noise processes

- A sequence of independently and identically distributed random variables $\{\varepsilon_t, t = 1, 2, ...\}$ with mean 0 and finite variance σ^2.

- Model
$$y_t = \varepsilon_t, \quad t = 1, 2, ...$$

- $\rho_k = \rho_{kk} = 0$, for all $k \geq 1$.

- Is this a stationary time series? Can you expect to capture any predictable pattern in this time series?

Outline

$AR(1)$ process

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t,$$

where ε_t is i.i.d. with mean zero and variance σ^2, i.e. $\{\varepsilon_t\}$ is a white noise process.

We will next find the mean, variance and covariance. First, the mean:

$$E(Y_t) = c + \phi_1 E(Y_{t-1}),$$

Under the assumption of stationarity $E(Y_t) = E(Y_{t-1})$, so

$$E(Y_t) = \frac{c}{1 - \phi_1}.$$

$AR(1)$ process: Properties

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t,$$

$$\mathbb{V}(Y_t) = \phi_1^2 \mathbb{V}(Y_{t-1}) + \sigma^2,$$

Under the assumption of stationarity $\mathbb{V}(Y_t) = \mathbb{V}(Y_{t-1})$, so

$$\mathbb{V}(Y_t) = \frac{\sigma^2}{1 - \phi_1^2}.$$

$AR(1)$ process: Properties

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t-1}) &= \mathrm{Cov}(c + \phi_1 Y_{t-1} + \varepsilon_t, Y_{t-1}) \\
&= \mathrm{Cov}(c, Y_{t-1}) + \mathrm{Cov}(\phi_1 Y_{t-1}, Y_{t-1}) + \mathrm{Cov}(\varepsilon_t, Y_{t-1}) \\
&= 0 + \phi_1 \mathrm{Var}(Y_{t-1}) + 0 = \phi_1 \mathrm{Var}(Y_{t-1}). \quad \text{Why?}
\end{aligned}
$$

$$
\rho_1 = \frac{\mathrm{Cov}(Y_t, Y_{t-1})}{\sqrt{\mathbb{V}(Y_t)\mathbb{V}(Y_{t-1})}} \overset{\text{Why?}}{=\!=} \frac{\mathrm{Cov}(Y_t, Y_{t-1})}{\mathbb{V}(Y_{t-1})} = \phi_1.
$$

$AR(1)$ process: Properties

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t-2}) &= \mathrm{Cov}(c + \phi_1 Y_{t-1} + \varepsilon_t, Y_{t-2}) \\
&= \mathrm{Cov}(\phi_1(c + \phi_1 Y_{t-2} + \varepsilon_{t-1}), Y_{t-2}) \\
&= \phi_1^2 \mathrm{Var}(Y_{t-2}).
\end{aligned}
$$

Thus, noting that $\mathrm{Var}(Y_{t-2}) = \mathrm{Var}(Y_{t-1}) = \mathrm{Var}(Y_t)$,

$$
\rho_2 = \frac{\mathrm{Cov}(Y_t, Y_{t-2})}{\mathbb{V}(Y_t)} = \phi_1^2,
$$

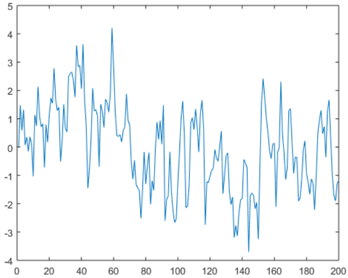
$$
\vdots \quad \text{(Similarly)}
$$

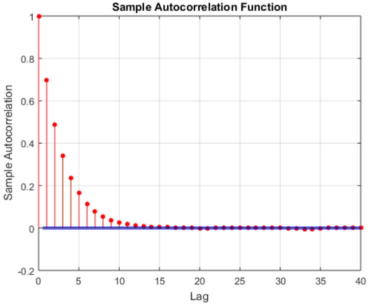
$$
\rho_k = \frac{\mathrm{Cov}(Y_t, Y_{t-k})}{\mathbb{V}(Y_t)} = \phi_1^k.
$$

$AR(1)$ process: Properties

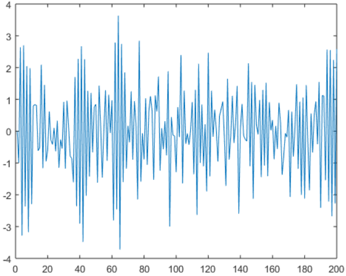
By the definition of Partial ACF, it's easy to see that: $\rho_{kk} = 0$ for all $k > 1$.

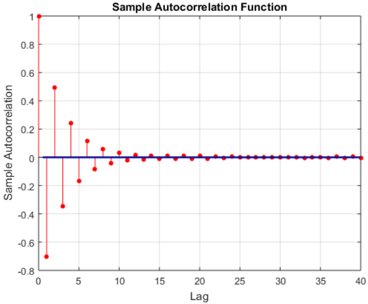
$AR(1)$ process: $\phi = 0.7$



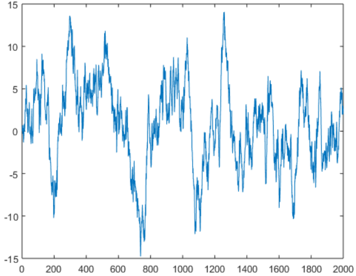


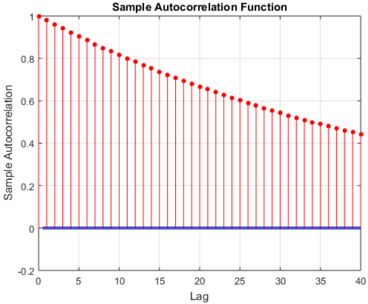
$AR(1)$ process: $\phi = -0.7$





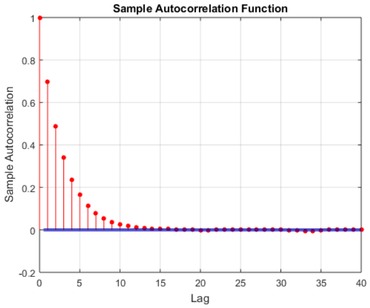
$AR(1)$ process: $\phi = 0.98$

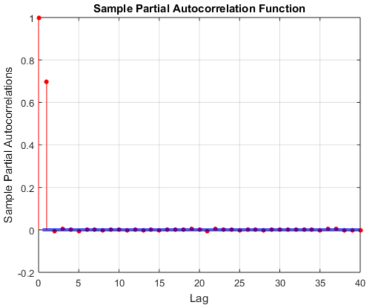




$AR(1)$ process

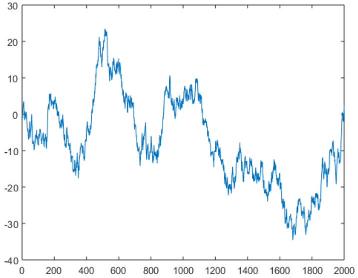
$\phi = 0.7$ ACF (left) and Partial ACF (right)

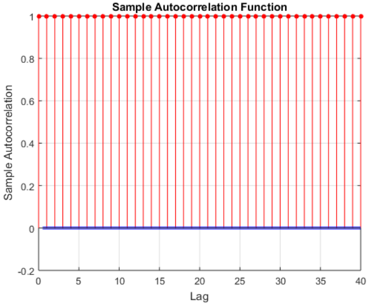




$AR(1)$ process: $\phi = 1$

What happens when $\phi_1 = 1$? See `Lecture07_Example02.py`





$AR(1)$ process: summary

- When $|\phi_1| < 1$, the AR(1) process is stationary
- ACF: $\rho_k = \phi_1^k$, $k = 0, 1, \ldots$
- Partial ACF: $\rho_{kk} = 0$ for all $k > 1$.
- How to check if a time series is an AR(1)?
  - The sample ACF plot dies down in a steady fashion
  - The sample Partial ACF cuts off after lag 1.

$AR(1)$ model: Forecasting

$$
\begin{aligned}
\widehat{y}_{t+1} &= \mathbb{E}(Y_{t+1}|y_{1:t}) \\
&= \mathbb{E}(c + \phi_1 Y_t + \varepsilon_{t+1}|y_{1:t}) \\
&= c + \phi_1 y_t.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}(Y_{t+1}|y_{1:t}) &= \mathbb{V}(c + \phi_1 Y_t + \varepsilon_{t+1}|y_{1:t}) \\
&= \mathbb{V}(c + \phi_1 y_t + \varepsilon_{t+1}|y_t) \\
&= \sigma^2.
\end{aligned}
$$

$AR(1)$ model: Forecasting

$$
\begin{aligned}
\widehat{y}_{t+2} &= \mathbb{E}(Y_{t+2}|y_{1:t}) \\
&= \mathbb{E}(c + \phi_1 Y_{t+1} + \varepsilon_{t+2}|y_{1:t}) \\
&= c + \phi_1 \mathbb{E}(Y_{t+1}|y_{1:t}) \\
&= c + \phi_1(c + \phi_1 y_t) \\
&= c(1 + \phi_1) + \phi_1^2 y_t.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}(Y_{t+2}|y_{1:t}) &= \mathbb{V}\big(c + \phi_1 Y_{t+1} + \varepsilon_{t+2}|y_{1:t}\big) \\
&= \phi_1^2 \mathbb{V}(Y_{t+1}|y_{1:t}) + \sigma^2 \\
&= (1 + \phi_1^2)\sigma^2.
\end{aligned}
$$

$AR(1)$ model: Forecasting

$$\begin{aligned}\widehat{Y}_{t+h} &= c + \phi_1 \widehat{Y}_{t+h-1} \\ &= c(1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^{h-1}) + \phi_1^h y_t\end{aligned}$$

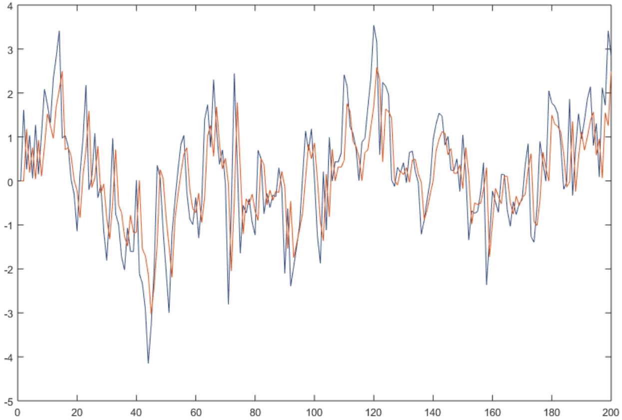
$$\begin{aligned}\mathbb{V}(Y_{t+h}|y_{1:t}) &= \phi_1^2 \mathbb{V}(Y_{t+h-1}|y_{1:t}) + \sigma^2 \\ &= \sigma^2(1 + \phi_1^2 + \ldots + \phi_1^{2(h-1)}).\end{aligned}$$

What happens as h gets larger?

$AR(1)$ process

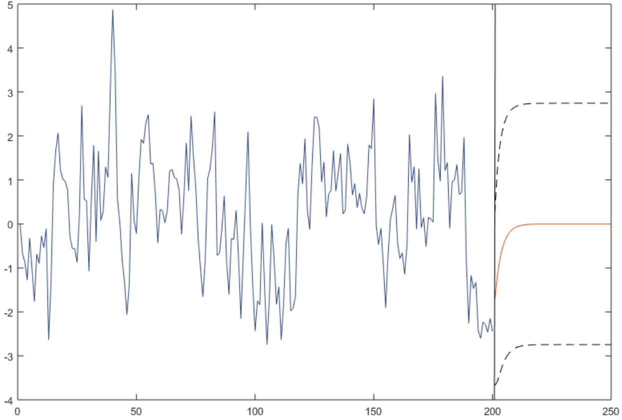
In-sample fit illustration

The red curve is $\widehat{y}_{t|t-1}$, $t = 1, ..., N$



$AR(1)$ process

Forecasting illustration



Outline

Stationarity, ACF and Partial ACF

Autoregressive process

- $AR(1)$ process
- $AR(p)$ process

$AR(p)$ processes: Properties

$$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t,$$

$$E(Y_t) = c + \phi_1 E(Y_{t-1}) + \ldots + \phi_p E(Y_{t-p})$$

Suppose it is stationary, then

$$\begin{aligned} E(Y_t) &= \frac{c}{1 - \phi_1 - \phi_2 - \ldots - \phi_p} \\ &= \frac{c}{1 - \sum_{i=1}^{p} \phi_i} \end{aligned}$$

$AR(p)$ processes: Properties

$$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t,$$

$$\mathbb{V}(Y_t) = \mathbb{V}(c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t)$$

Can we continue like this?

$$\mathbb{V}(Y_t) = \mathbb{V}(c) + \mathbb{V}(\phi_1 Y_{t-1}) + \ldots + \mathbb{V}(\phi_p Y_{t-p}) + \mathbb{V}(\varepsilon_t)$$

NO! because all

$$\mathrm{Cov}(Y_{t-1}, Y_{t-2}) \neq 0$$

Under the stationary condition, it can be proved that

$$\mathbb{V}(Y_t) = \frac{\sigma^2}{(1-\rho_{11}^2)(1-\rho_{22}^2)\ldots(1-\rho_{pp}^2)}$$

$AR(2)$ processes: Properties

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t-1}) &= \mathrm{Cov}(c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, Y_{t-1}) \\
&= \phi_1 \mathbb{V}(Y_{t-1}) + \phi_2 \mathrm{Cov}(Y_{t-2}, Y_{t-1})
\end{aligned}
$$

Under the stationary condition we have

$$
\mathrm{Cov}(Y_t, Y_{t-1}) = \mathrm{Cov}(Y_{t-2}, Y_{t-1}) = \frac{\phi_1}{1-\phi_2} \mathbb{V}(Y_{t-1}).
$$

$$
\rho_1 = \frac{\mathrm{Cov}(Y_t, Y_{t-1})}{\sqrt{\mathbb{V}(Y_t)\mathbb{V}(Y_{t-1})}} = \frac{\phi_1}{1-\phi_2}.
$$

where we have used $\mathbb{V}(Y_t) = \mathbb{V}(Y_{t-1})$.

$AR(2)$ processes: Properties

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t-2}) &= \mathrm{Cov}(c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, Y_{t-2}) \\
&= \phi_2 \mathbb{V}(Y_{t-2}) + \phi_1 \mathrm{Cov}(Y_{t-1}, Y_{t-2}) \\
&= \left(\phi_2 + \frac{\phi_1^2}{1-\phi_2}\right) \mathrm{Var}(Y_{t-2}).
\end{aligned}
$$

$$
\rho_2 = \frac{\mathrm{Cov}(Y_t, Y_{t-2})}{\sqrt{\mathbb{V}(Y_t)\mathbb{V}(Y_{t-2})}} = \phi_2 + \frac{\phi_1^2}{1-\phi_2}.
$$

where we have used $\mathbb{V}(Y_t) = \mathbb{V}(Y_{t-2})$.

$AR(2)$ processes: Properties

$$\begin{aligned}
\text{Cov}(Y_t, Y_{t-3}) &= \text{Cov}(c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, Y_{t-3}) \\
&= \phi_1 \text{Cov}(Y_{t-1}, Y_{t-3}) + \phi_2 \text{Cov}(Y_{t-2}, Y_{t-3}) \\
&= \phi_1 \rho_2 \mathbb{V}(Y_{t-3}) + \phi_2 \rho_1 \mathbb{V}(Y_{t-3}).
\end{aligned}$$

where we have used $\rho_2 = \frac{\text{Cov}(Y_{t-1}, Y_{t-3})}{\mathbb{V}(Y_{t-3})}$ and $\rho_1 = \frac{\text{Cov}(Y_{t-2}, Y_{t-3})}{\mathbb{V}(Y_{t-3})}$.

$$\rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1$$

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k > 2$$

$AR(p)$ processes: Properties

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t,$$

where ε_t is i.i.d. with mean zero and variance σ^2.
It can be shown that

- ACF ρ_k dies down exponentially.
- PACF ρ_{kk} cuts off to zero after lag p.

These properties are useful to recognize an AR(p) process.

$AR(p)$ processes: Forecasting

$$\widehat{y}_{t+h} = E(Y_{t+h}|y_{1:t}) = c + \phi_1 E(Y_{t+h-1}|y_{1:t}) + \ldots + \phi_p E(Y_{t+h-p}|y_{1:t}),$$

where

$$E(Y_{t+h-i}|y_{1:t}) = \begin{cases} \widehat{y}_{t+h-i} & \text{if } h > i \\ y_{t+h-i} & \text{if } h \leq i. \end{cases}$$

For example, consider AR(3),

$$Y_{t+1} = c + \phi_1 Y_t + \phi_2 Y_{t-1} + \phi_3 Y_{t-2} + \varepsilon_{t+1}$$

then

$$\begin{aligned} \widehat{y}_{t+1} &= c + \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2} \\ \widehat{y}_{t+2} &= c + \phi_1 \widehat{y}_{t+1} + \phi_2 y_t + \phi_3 y_{t-1} \\ \widehat{y}_{t+3} &= c + \phi_1 \widehat{y}_{t+2} + \phi_2 \widehat{y}_{t+1} + \phi_3 y_t \end{aligned}$$

$AR(p)$ processes: Forecasting

Hence

$$
\begin{aligned}
\widehat{y}_{t+1} =& c + \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2} \\
\widehat{y}_{t+2} =& c + \phi_1 \widehat{y}_{t+1} + \phi_2 y_t + \phi_3 y_{t-1} \\
=& c + \phi_1 (c + \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2}) + \phi_2 y_t + \phi_3 y_{t-1} \\
=& c(1 + \phi_1) + (\phi_1^2 + \phi_2) y_t + (\phi_1 \phi_2 + \phi_3) y_{t-1} + \phi_1 \phi_3 y_{t-2} \\
\widehat{y}_{t+3} =& c + \phi_1 \widehat{y}_{t+2} + \phi_2 \widehat{y}_{t+1} + \phi_3 y_t \\
=& \cdots\cdots
\end{aligned}
$$

Finally what about the variance?

Recap

We have looked at

- Stationarity, ACF and Partial ACF
- Autoregressive processes: some properties and how to perform forecasting

Next lecture: Moving Average processes and ARIMA

Thank you and good luck with your midterm exam!